McGraw-Hill Encyclopedia of Science and Technology

McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON



of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 9 NAI-PEP

(LEFT) High-altitude infrared oblique photograph of Long Island, N. Y. *(Aero Service Corp.)*. (RIGHT) Large-scale photograph of the chromosphere of the Sun in Hα light, with the disk of the Sun artificially eclipsed *(photograph by R. B. Dunn using 15-in tele-scope, Sacramento Peak Observatory)*.

Library of Congress Catalog Card Number: 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use

The subject matter of the various disciplines or branches of science and technology is organized systematically a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject The cross references—there are about 50 000 of them—are printed in small capital letters so that they can be easily recognized By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on

In general, each article begins with a definition of the title that states its scope and coverage. Usually, only the scientific or technological sense is discussed Most of the articles, after this statement, go on to increasingly complex and detailed considerations A reader thus needs to proceed only as far as his inclinations and requirements dictate

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries.

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name

The adjectives electric and electrical are used in the following senses Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter; for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally, thus.

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words; for example,

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article) Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles

A list of initials and names of the contributors to the Encycloped[illegible] [illegible] encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia

McGraw-Hill Encyclopedia of Science and Technology

N Nailing to Nystatin

Nailing

The driving of nails in a manner that will position and hold two or more members, usually of wood, in a desired relationship to each other The contact pressures between the surfaces of the nails and the surrounding wood fibers hold the nails in position

Strength of a nailed joint. Factors that determine the strength and efficiency of a nailed joint are (1) the type of wood, (2) the nail used, (3) the conditions under which the nailed joint is used; and (4) the number of nails

In general, hard, dense woods hold nails better than soft woods The better the resistance of a nail to direct withdrawal from a piece of wood, the tighter the joint will remain. Nails driven into green wood tend to loosen slightly as the wood dries and shrinks In seasoned material the resistance to withdrawal diminishes only slightly with time, unless moisture affects the wood Withdrawal resistance is always higher when nails are driven into the side grain than when into the end grain. Because the lighter woods do not usually split as readily as the denser ones, more and larger nails can be used to offset the poorer nail holding properties of the former Hardwoods are more difficult to nail, they are sometimes used green or with holes drilled for nailing, to prevent splitting

The surface condition of a nail affects its holding ability The withdrawal resistance of a common nail increases directly with the distance it penetrates into the wood and increases almost directly with its surface area A rusty nail may offer more resistance to withdrawal than a smooth one

Means to increase withdrawal resistance. To increase resistance to withdrawal or loosening, nails may be coated, etched, spirally grooved, annularly grooved, or barbed, as illustrated Grooved nails tend to hold well despite a change in moisture content Coated nails usually provide a greater increase in withdrawal resistance when used in softer woods than when used in the denser woods. The increase in withdrawal resistance tends to decrease, however with time

In most cases, nails driven on a slant have more withdrawal resistance than nails driven straight into the wood If a slant driven nail is pulled in a direction which is at right angles to the surface considerable resistance is encountered from the wood fibers on the pressure side. The nail may also progressively bend as it is pulled out Both of these factors seem to offer continued holding power, even though the wood fibers are not gripping the entire surface of the nail Nails slant-driven into the end grain of wood seem to gain proportionately more withdrawal resistance than those slant-driven into the side grain.

When members of a nailed joint tend to separate sideways, the nails are subjected to side loads. In this case doubling the diameter of the nail increases its lateral load capacity by nearly three times. This is true, however, only if the nail point has been driven a suitable distance into the piece receiving it.

Blunt pointed nails are of...

Nail	
spiral-threaded insulated siding face nail	
annular-ring gypsum board drywall nail	
asbestos shingle nails — annular-ring	
asbestos shingle nails — spiral-threaded	
annular-ring plywood roofing nail for applying wood or asphalt shingles over plywood sheathing	
annular-ring plywood siding nail for applying asbestos shingles and shakes over plywood sheathing	
spiral-threaded casing head wood siding nail	
annular-ring roofing nail for asphalt shingles and shakes	
spiral-threaded roofing nail for asphalt shingles and shakes	
annular-ring roofing nail with neoprene washer	
spiral-threaded roofing nail with neoprene washer	
insulated siding nail	
gypsum lath nail	
wood shake nail	
wood shingle nail	
roofing nail	
general purpose finish nail	
sinker head wood siding nail	
casing head wood siding nail	

Special- and general-purpose nails

greater number of nails per joint. Beeswax is sometimes applied to nail points to make them drive more easily, but it also reduces the holding power of the nail. *See* WOODWORKING. [A.T.]

Naphtha

Any one of a wide variety of volatile hydrocarbon mixtures. They are sometimes obtained from coal tar but are more often derived from petroleum. Physical properties vary widely. The initial boiling point may be as low as 80°F, and end points may reach 500°F. Boiling ranges are sometimes as narrow as 20° or as wide as 200°.

The main process for producing naphthas is fractional distillation. It may be of the extractive type when certain high-quality naphthas are desired. Acid treating, clay treating, and other techniques remove sulfur compounds and improve color, odor, and stability.

Strictly speaking, the refinery streams going into products like gasoline and kerosine are naphthas, and they are so designated within the petroleum industry. The final blended fuels, however, are sold under the more familiar names. The products sold as naphthas find their greatest use as solvents, thinners, or carriers.

Few naphthas are made up entirely of hydrocarbons belonging to one particular family. There is a fairly sharp differentiation, however, between aliphatic and aromatic types.

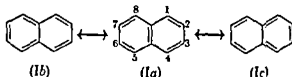
Aliphatic naphthas are relatively low in odor and toxicity and tend, also, to be low in solvent power, which in some cases is an advantage. In the processing of soybeans, for example, the aim is to extract the oil without extracting the less desirable materials. Naphthas used by dry cleaners likewise require only moderate solvent power. In printing ink, the naphtha is mainly a carrier of the carbon black or other pigments; resins requiring a solvent are present in only minor amounts.

The aromatic naphthas, often described as the "high-solvency" type, at one time came entirely from coal tar. The development of catalytic cracking and catalytic reforming made petroleum an alternative source. The main components are toluene and xylenes; benzene is less desirable because of the extreme toxicity of its vapors. A major use of these naphthas is as thinners for paints and varnishes, to permit easy brushing. Both varnishes and enamels contain large amounts of gums and resins, and diluents with good solvent action are therefore needed.

The rubber industry also uses naphthas as solvents. The leather industry uses them to degrease skins, the metal industry to degrease metals. Naphthas in insecticides and weed-killers dissolve the toxic agents and often contribute toxic properties of their own. Floor waxes, furniture waxes, shoe polishes, metal polishes, and dry cleaners' soaps are among the many other products in which naphthas are used. *See* PETROLEUM PRODUCTS; PETROLEUM REFINING. [J.K.R.]

Naphthalene

A colorless crystalline aromatic hydrocarbon ($C_{10}H_8$) with the familiar odor of moth balls, melting point 80.1°, boiling point 218°C. It is almost insoluble in water but soluble in nearly all organic solvents. Structurally it is best represented as two benzenoid rings fused together (*Ia*). In the

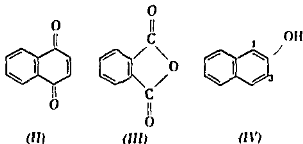


(*Ib*) (*Ia*) (*Ic*)

naming of naphthalene derivatives, numbers are most frequently used (*Ia*) especially for di- and polysubstituted compounds, but the use of α for position 1 and β for position 2 is still encountered.

Occurrence. One gallon of coal tar yields approximately one pound of naphthalene. It is present also in certain petroleum fractions which have been subjected to cracking, but separation of the naphthalene is not economically justified at present. Crude naphthalene contains a small quantity of sulfur which can be removed by distilling the crude hydrocarbon from sodium metal. Thisnaphthalene (C_8H_6S) may be removed from naphthalene by preferential chlorination.

Reactions. Naphthalene is considered less aromatic than benzene since it is more easily reduced and oxidized, and it shows a greater tendency to react by addition. Substitution reactions occur much more rapidly with naphthalene than with benzene. With sodium in boiling absolute ethanol, it yields 1,4-dihydronaphthalene, while with sodium in boiling amyl alcohol, it affords 1,2,3,4-tetrahydronaphthalene (tetralin). Oxidation of naphthalene with chromic acid yields some 1,4-naphthoquinone (II) while vapor-phase air oxida-



(*II*) (*III*) (*IV*)

tion over a vanadium pentoxide catalyst yields phthalic anhydride (III), which is important in the manufacture of glyptal resins.

The α positions of the naphthalene nucleus are more reactive than the corresponding β positions. Nitration yields 1-nitronaphthalene almost free of the 2-nitro isomer. Halogenation in the presence of a catalyst yields the 1-halonaphthalene. Without a catalyst, chlorination occurs by addition, yielding 1,2,3,4-tetrachloro-1,2,3,4-tetrahydronaphthalene. Sulfonation with concentrated sulfuric acid at low temperatures yields naphthalene-1-sulfonic acid, while at high temperatures, naphthalene-2

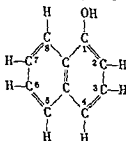
sulfonic acid ($C_{10}H_7SO_3H$) is obtained. The Friedel Crafts reaction occurs readily, usually affording mixtures of 1- and 2-substituted naphthalenes

Bond structure. Naphthalene has been stated to be a resonance hybrid of 42 canonical forms and of these, the three unexcited states are represented by (I*a*), (I*b*), and (I*c*) The Erlenmeyer formula (I*a*) is commonly used for naphthalene since it clearly indicates which of the peripheral bonds have the greatest double-bond character As would be predicted from the Erlenmeyer formula, 2-naphthol (IV) is activated at position 1, but not at position 3

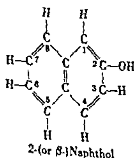
Uses. In 1957, 513,000,000 lb of naphthalene was consumed, with 82% being converted to phthalic anhydride. Aside from 2% used in the manufacture of moth balls, the remainder was converted to naphthalene compounds which are used as dye intermediates, tanning agents, and surface-active agents. *See* AROMATIC HYDROCARBON, POLYNUCLEAR HYDROCARBON. [C.K.B.]

Naphthol

One of those phenols that have a hydroxyl group bound directly to a naphthalene ring system of carbon atoms Two simple naphthols are known: 1-(or α-)naphthol and 2-(or β-)naphthol. The for-



1-(or α-)Naphthol

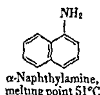


2-(or β-)Naphthol

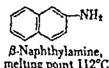
mulas show the numbers assigned to the carbon atoms of the *ring system* These numbers are used to locate other atoms or atomic groups that are present in more complex naphthols. Both α- and β-naphthols are produced by fusion of the corresponding naphthalene sulfonates with caustic soda. The naphthols are used as intermediates from which dyes are made. The demand is chiefly for β-naphthol, so that about 50 times more β- than α-naphthol is produced commercially. *See* PHENOL. [R B C.]

Naphthylamine

One of two organic *chemical compounds*, nearly insoluble in water, that are used to make various sulfonic acid derivatives to serve as coupling components for azo dye intermediates.



α-Naphthylamine, melting point 51°C



β-Naphthylamine, melting point 112°C

α-Naphthylamine is obtained by the reduction of α-nitronaphthalene. β-Naphthylamine is manufactured by heating β-naphthol in an autoclave with a solution of ammonia and ammonium sulfite (Bucherer process). The naphthylamines resemble aniline in their properties.

Heating the sulfate salt of α-naphthylamine yields 1-naphthylamine-4-sulfonic acid (*naphthionic acid*) in analogy to aniline sulfate (*see* SULFANILIC ACID). Bucherer's reaction with sodium hydrogen sulfite and sodium hydroxide leads to 1-naphthol-4-sulfonic acid (Neville-Winter acid). By sulfonation of naphthalene, followed by nitration and reduction of the [illegible]

[illegible]

From β-naphthylamine by various sulfonation and hydrolysis reactions, the following dye intermediates result: 2-naphthylamine-1-sulfonic acid (Tobin's acid), 2 naphthylamine-6,8-disulfonic acid, 2-naphthylamine-1,5,7-trisulfonic acid. 2- [illegible]

[illegible]

ge [illegible]

Ar [illegible]

[illegible] L. F Fieser and M. Fieser, *Organic Chemistry*, 3d ed., 1956; H A. Lubs, *The Chemistry of Synthetic Dyes and Pigments*, 1955; K. Venkataraman, *The Chemistry of Synthetic Dyes*, 2 vols., 1952.

Narcotic

Any drug which will induce sleep or coma and which will alleviate pain. The oldest of these are probably opium and some of its derivatives, originating from the dried juices of the poppy seed. Other derivatives in common use include morphine, paregoric, and codeine. In the United States, the Harrison Narcotics Law and its amendments [illegible] late the [illegible]

opiu [illegible]

tives [illegible]

cons[illegible] under the Harrison Act follows. [illegible] of drugs

- Extract of opium
- Powdered opium
- Ipecac and opium
- Tincture of opium
- Morphine
- *Magendie's solution*
- Codeine
- Codeonal
- Heroin
- Dionine
- Pantopon
- Dilaudid
- Narcofine
- Papaverine
- Apomorphine
- *Meperidine*
- Methadone
- Stypticin
- Styptol
- Demerol
- Bemidone
- Metopon
- Nisentil
- Cocaine
- Tropacaine
- *Holacaine*
- Eucaine
- Dromoran

New synthetics are added to the list as they appear In addition to several Federal laws, many states have passed legislation to cover particular drugs, or to extend the regulation under Federal controls These were especially aimed at the barbiturates and marijuana, both of which present problems in abuse, particularly in large metropolitan areas.

It is estimated that there are between 60,000 and 100,000 drug addicts in the United States, although the basis for such figures is open to question.

Although a relatively small percentage of persons become addicted as a result of prolonged drug therapy during illnesses, by far the greater proportion become addicted through illicit means. Addiction implies a defective personality in the vast majority of cases, so that an insecure individual turns to drugs to alleviate real or imagined conditions of psychic stress.

The continued use of narcotics is marked by an increasing tolerance to their effects and by physical dependence upon the drugs to prevent the symptoms of withdrawal The tolerance often produces fantastic requirements of a particular drug in order for the desired sensations to be elicited The social significance lies not only in the drug addiction itself, but in the increasing costs of the habit. Invariably, normal employment will not support these costs and therefore some form of crime is resorted to in an attempt to meet the high costs.

The symptoms of chronic drug addiction may be illustrated by those produced by morphine Nausea, vomiting, sweating, itching, and pallor are commonplace in the early stages. There is a decrease in appetite and, despite common belief, the sexual drive and ability are lessened. Mental states of confusion, disorientation, and hallucinations may occur, and daydreaming or fantasy is commonplace. Chronic constipation, with or without bouts of diarrhea, constriction of the pupils, and vasomotor reactions mark the progress of addiction.

Withdrawal symptoms are the dreaded result of inability to obtain more narcotic. The patient's entire body must adjust to the deprivation Acute depression, insomnia, severe cramps and muscular pains, trembling, extreme sweating, and many other general and local symptoms appear.

Treatment for drug addiction requires a great deal of specialized attention. Psychotherapy must be combined with substitution of less-toxic preparations so as to minimize the withdrawal symptoms, although many authorities believe in the efficacy of abrupt withdrawal of all narcotics. Follow-up of cases is most important, since psychic factors are ordinarily at the root of the individual's addiction In the United States, two Federal sanitariums for drug addicts are maintained, at Lexington, Kentucky, and Fort Worth, Texas. In addition, there are numerous state and private institutions which are equipped to treat the narcotics addict. *See* HEADACHE; PAIN, CUTANEOUS; PAIN, DEEP.

[F.G.ST.]

Native elements

Those elements which occur in nature uncombined with other elements Aside from the free gases of the atmosphere there are about 20 elements that are found as minerals in the native state. These are divided into metals, semimetals and nonmetals. Gold, silver, copper, and platinum are the most important metals and each of these has been found abundantly enough at certain localities to be mined as an ore Native gold and platinum are the major ore minerals of these metals Rarer native metals are others of the platinum group, lead, mercury, tantalum, tin, and zinc Native iron is found sparingly both as terrestrial iron and meteoric iron. *See* ORE AND MINERAL DEPOSITS.

The native semimetals can be divided into (1) the arsenic group, including arsenic, antimony and bismuth, and (2) the tellurium group, including tellurium and selenium. The members of the arsenic group crystallize in the hexagonal system, scalenohedral class; those of the tellurium group in the hexagonal system, trigonal trapezohedral class. Only rarely do the semimetals occur abundantly enough to be mined as ores of their respective elements *See* MINERALOGY.

The native nonmetals are sulfur, and carbon in the forms of graphite and diamond Native sulfur is the chief industrial source of that element

[C.S.HU.]

Natrolite

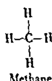
A fibrous or needlelike mineral belonging to the zeolite family of silicates It crystallizes in the monoclinic system in pseudo-orthorhombic prismatic crystals which are often acicular Most commonly it is found in radiating fibrous aggregates There is perfect prismatic cleavage, the hardness is 5-5½ on Mohs scale, and the specific gravity is 2.25 The mineral is white or colorless with a vitreous luster that inclines to pearly in fibrous varieties The chemical composition is $Na_2(Al_2Si_3O_{10}) \cdot 2H_2O$ but some potassium is usually present substituting for sodium

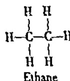
Natrolite is a secondary mineral (low temperature hydrothermal mineral) found lining cavities in basaltic rocks, where it is associated with other zeolites, calcite, apophyllite, and prehnite Its outstanding locality in the United States is at Bergen Hill, New Jersey *See* ZEOLITE. [C.F.B., C.S.HU.]

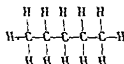
Natural gas

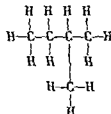
Inflammable gas that occurs in porous rock of the earth's crust and is found with or near accumulations of crude oil. Being in gaseous form, it may occur alone in separate reservoirs. More commonly it forms a gas cap or mass of gas entrapped between liquid petroleum and impervious capping rock layer in a petroleum reservoir. Under conditions of greater pressure it is intimately mixed with, or dissolved in, crude oil

Composition. *Typical natural gas consists of* hydrocarbons having a very low boiling point Methane (CH_4), the fundamental member of the methane series, with a boiling point of −254°F, makes up approximately 85% of the typical gas. Ethane (C_2H_6), with a boiling point of −128°F, is likely to be present in amounts up to 10%; and propane (C_3H_8), with a boiling point of −44°F, up to 3%. Butane (C_4H_{10}), pentane (C_5H_{12}), hexane, heptane, and octane may be present.


Methane


Ethane


n-Pentane


i-Pentane

Whereas normal hydrocarbons having 5–10 carbon atoms are liquids at ordinary temperatures, these paraffins of higher molecular weight are present in vapor form Impurities present in considerable amounts are carbon dioxide, nitrogen, helium, and hydrogen sulfide

Types of natural gas vary according to composition to result in a dry or lean (mostly methane) gas wet gas (considerable amounts of so called higher hydrocarbons), sour gas (much hydrogen sulfide) sweet gas (little hydrogen sulfide), residue gas (higher paraffins having been extracted), and casinghead gas (derived from an oil well by extraction at the surface) Nearly all natural gas is inflammable It has no distinct odor Its main use is for fuel but it is also used to make carbon black, natural gasoline, liquefied petrole [illegible] tain che [illegible]

Distri [illegible]

contin [illegible]

amoun [illegible]

[illegible] reserves (Texas, Louisiana, New Mexico, Kansas, Oklahoma, and California). Among these Texas has 47% of the total with 115×10^{12} ft³, whereas Louisiana ranks second with 19% or 46×10^{12} ft³. The estimated known reserves in the United States at the end of 1957 were 240×10^{12} ft³. Consumption in the United States in 1957 was 10,279,775,000,000 cubic feet. The annual rate of finding new reserves in the United States is about 12×10^{12} ft³. New reserves are being discovered in western Canada at a rapid rate and the total ultimately will be very great.

In estimating gas reserves the volumetric method is preferred The volume of the reservoir is determined by means of the thickness, porosity, and permeability of the producing zones. A study of many depleted fields suggests that about 85% of all gas in dry-gas reservoirs is recovered. Some engineers use the production versus pressure decline method. They calculate future production by plotting past production against the decline in reservoir pressure.

In California 75% of the gas is associated with oil, but in West Texas the percentage is even higher. In southern Texas the percentage is also high By contrast, the percentage figure for the United States as a whole is only 30%. This means that a large proportion of the reserves is stored in such dry-gas fields as the Hugoton (in Kansas and adjacent Oklahoma and Texas); the Monroe (in Louisiana); and the Carthage (in northeastern Texas) In western Canada some of the large gas pools are the Pincher Creek, the Waterton, and the *Jumping Pound* pools The largest dissolved-gas area in the United States lies along the Gulf Coast of Texas and Louisiana It contains about 35% of the total known reserves. Offshore *drilling in the* waters of the Gulf will add considerably to these reserves *In an* average year, slightly over 97% of the gas produced is marketed, while 1.5% is used for [illegible]

[illegible] system of rocks down to the Cambrian The first gas deposits found in the United States were those in the eastern states. In New York and Pennsylvania 85% of the gas came from Devonian rocks In West Virginia, Kentucky, and eastern Ohio, Devonian and Mississippian rocks rank nearly equal, but Silurian rocks are also important. In Indiana and Illinois, Pennsylvanian rocks outrank the Mississippian. The Hugoton field in Kansas is one of the largest in the world Here Permian dolomites produce gas from five different levels. The fact that oil is found lower down in Pennsylvanian and older rocks proves the superior migratory capacity of gas. Up to the end of 1957 the Kansas portion of this field had produced about 4.5×10^{12} ft³. One notable fact about the field is the high percentage of nitrogen (almost 15%). The Hugoton producing area extends across the Oklahoma Panhandle and almost across the Texas Pan-

Extract of opium	Pantopon	Bemidone
Powdered opium	Dilaudid	Metopon
Ipecac and opium	Narcofine	Nisentil
Tincture of opium	Papaverine	Cocaine
Morphine	Apomorphine	Tropacaine
Magendie's solution	Meperidine	Holacaine
Codeine	Methadone	Eucaine
Codeonal	Stypticin	Dromoran
Heroin	Styptol	
Dionine	Demerol	

New synthetics are added to the list as they appear. In addition to several Federal laws, many states have passed legislation to cover particular drugs, or to extend the regulation under Federal controls. These were especially aimed at the barbiturates and marijuana, both of which present problems in abuse, particularly in large metropolitan areas.

It is estimated that there are between 60,000 and 100,000 drug addicts in the United States, although the basis for such figures is open to question.

Although a relatively small percentage of persons become addicted as a result of prolonged drug therapy during illnesses, by far the greater proportion become addicted through illicit means. Addiction implies a defective personality in the vast majority of cases, so that an insecure individual turns to drugs to alleviate real or imagined conditions of psychic stress.

The continued use of narcotics is marked by an increasing tolerance to their effects and by physical dependence upon the drugs to prevent the symptoms of withdrawal. The tolerance often produces fantastic requirements of a particular drug in order for the desired sensations to be elicited. The social significance lies not only in the drug addiction itself, but in the increasing costs of the habit Invariably, normal employment will not support these costs and therefore some form of crime is resorted to in an attempt to meet the high costs.

The symptoms of chronic drug addiction may be illustrated by those produced by morphine. Nausea, vomiting, sweating, itching, and pallor are commonplace in the early stages. There is a decrease in appetite and, despite common belief, the sexual drive and ability are lessened. Mental states of confusion, disorientation, and hallucinations may occur, and daydreaming or fantasy is commonplace. Chronic constipation, with or without bouts of diarrhea, constriction of the pupils, and vasomotor reactions mark the progress of addiction.

Withdrawal symptoms are the dreaded result of inability to obtain more narcotic. The patient's entire body must adjust to the deprivation. Acute depression, insomnia, severe cramps and muscular pains, trembling, extreme sweating, and many other general and local symptoms appear.

Treatment for drug addiction requires a great deal of specialized attention. Psychotherapy must be combined with substitution of less-toxic preparations so as to minimize the withdrawal symptoms, although many authorities believe in the efficacy of abrupt withdrawal of all narcotics. Follow-up of cases is most important, since psychic factors are ordinarily at the root of the individual's addiction. In the United States, two Federal sanitariums for drug addicts are maintained, at Lexington, Kentucky, and Fort Worth, Texas. In addition, there are numerous state and private institutions which are equipped to treat the narcotics addict. See HEADACHE; PAIN, CUTANEOUS; PAIN, DEEP.

[E.C.ST.]

Native elements

Those elements which occur in nature uncombined with other elements. Aside from the free gases of the atmosphere there are about 20 elements that are found as minerals in the native state. These are divided into metals, semimetals and nonmetals. Gold, silver, copper, and platinum are the most important metals and each of these has been found abundantly enough at certain localities to be mined as an ore. Native gold and platinum are the major ore minerals of these metals. Rarer native metals are others of the platinum group, lead, mercury, tantalum, tin, and zinc. Native iron is found sparingly both as terrestrial iron and meteoric iron. See ORE AND MINERAL DEPOSITS

The native semimetals can be divided into (1) the arsenic group, including arsenic, antimony and bismuth; and (2) the tellurium group, including tellurium and selenium The members of the arsenic group crystallize in the hexagonal system, scalenohedral class; those of the tellurium group in the hexagonal system, trigonal trapezohedral class Only rarely do the semimetals occur abundantly enough to be mined as ores of their respective elements. See MINERALOGY.

The native nonmetals are sulfur, and carbon in the forms of graphite and diamond Native sulfur is the chief industrial source of that element

[C.S.HU.]

Natrolite

A fibrous or needlelike mineral belonging to the zeolite family of silicates It crystallizes in the monoclinic system in pseudo-orthorhombic prismatic crystals which are often acicular. Most commonly it is found in radiating fibrous aggregates There is perfect prismatic cleavage, the hardness is 5–5½ on Mohs scale, and the specific gravity is 2.25. The mineral is white or colorless with a vitreous luster that inclines to pearly in fibrous varieties. The chemical composition is $Na_2(Al_2Si_3O_{10})\cdot 2H_2O$ but some potassium is usually present substituting for sodium

Natrolite is a secondary mineral (low-temperature hydrothermal mineral) found lining cavities in basaltic rocks, where it is associated with other zeolites, calcite, apophyllite, and prehnite. Its outstanding locality in the United States is at Bergen Hill, New Jersey. See ZEOLITE. [C.FR.; C.S.HU.]

The chambered nautilus is one of the best known and most unusual of marine animals. The shell of this animal is highly prized as an ornament and souvenir. The shells are also in some demand for the high quality of their mother-of-pearl, and some of the finest cameos are cut from nautilus shells In the Philippines, where the animal abounds, it is used for food.

This species and the argonaut, or paper nautilus, are the only two cephalopods with an external shell; however, their shells are quite different in nature. The shell of the chambered nautilus is a flat spiral, yellowish in color, and marked with cross bands of brown. Inside it is brilliant pearl Removal of the outer layer with acid reveals the rest of the shell to be pearly throughout. In the adult the shell has two and one-half coils. Cross septa (walls) mark the shell cavity off into chambers of gradually increasing size, but all except the outer one are relatively small. The adult measures 4–6 in in diameter.

The animal lives in the outer chamber. The adult is about the size of a man's fist. There are two strong jaws on the conical head and about 100 tentacles arranged in four groups The eyes are dark, simple, large, and lateral The animal is white with some brown and yellow markings Its anatomy is much like that of the squid

The chambered nautilus is found throughout the oceanic depths, but is most abundant in the tropical seas bordering the Fiji Islands, the Philippines, New Hebrides, and New Caledonia.

Contrary to popular opinion, the nautilus does not normally float at the [illegible]

[illegible] the surface are usually dead or dying They are readily trapped in bamboo traps baited with crab meat and set on the bottom of the sea Details of their life history are unknown.

The argonaut, *Argonauta argo*, of the family Argonautidae, is often called the paper nautilus, although it is related only distantly to the true nautilus It is 4–8 in. across, with a double-keeled, ship-shaped, thin shell which is porcellanous and whitish tinged with yellow This shell is not secreted by the mantle, as are those of all other shelled Mollusca, but rather by the specialized, broad ends of the two dorsal arms. It is not otherwise attached to the body, and thus is not a true shell, but an egg cradle Only the females are shelled The males are virtually typical octopi, about an inch long, with eight arms, one of which is specialized to carry the sperm When sperm is mature, mating consists of a brief union of the male and female at which time the sperm-filled arm is broken off and left in the mantle of the female The egg mass, appearing like a cluster of grapes, remains in the cradle until the hatching young swim away

The argonaut, named after the Argonauts who, with Jason, sought the fabled Golden Fleece, lives in all tropical and warm seas The family Argonautidae is now placed in the suborder Octopoda There are five living species, but only *Argonauta argo* is well known. *See* CEPHALOPODA; OCTOPUS; SQUID. [J.D B.]

Naval architecture

The science that determines the physical characteristics of buoyant structures which operate in water. These structures are classified as ships, boats, submarines, barges, floats, and so forth, depending upon their intended service.

Naval architecture has lagged behind other sciences because it is not an exact science susceptible to precise mathematical treatment. It should actually be considered an art with scientific foundations rather than as a science. Much of naval architecture is based on a great fund of knowledge which has been accumulated from ages of practical experience. This knowledge is utilized mainly as an art rather than a science since many of the problems concerning the performance of ships are decided empirically from experience, rather than by precise scientific information

Naval architecture generally pertains to the design and construction of ships, both large and small, which meet military, commercial, and social requirements It is concerned with the ship as a whole rather than with a single specific part which is carried in a ship, either as cargo or a component such as the propulsion machinery, and includes the whole problem of the size, form, power, and strength of the ship. It includes also the economics and efficiency of performance, the determination of suitable ship dimensions, and the safety and comfort of the crew and passengers

Naval architecture requires that the designer possess creative imagination as well as technical skill. The naval architect is responsible for the ship as a complete entity and should therefore be familiar with all that pertains to a ship This requires knowledge concerning marine engineering; speed and propulsion; electrical engineering; hydrodynamics, the loading of cargo ships; the regulations of the various governmental agencies concerned with the operation and safety of ships, the classification societies concerned with the strength and insurability of ships and their cargo, structure, freeboard, and equipment; and a thorough knowledge of the practical work of the shipyard *See* DRYDOCKING; HYDROFOIL CRAFT; INLAND WATERWAYS TRANSPORTATION; LANDING SHIPS AND CRAFT; SHIP, MERCHANT; SHIP, NAVAL; SHIP DESIGN, SHIPBUILDING; SUBMARINE, *see also* MARINE ENGINEERING. [J.C N.]

Naval meteorology

The study of meteorology as it applies to operations at sea. Developments and contributions in this field have led to several outstanding achievements, among them wave forecasting, more efficient use of radar and sonar at sea, better forecasting of

handle, but the Kansas portion contains over 50% of the available gas.

The state of Oklahoma and the western part of Texas have gas in many stratigraphic zones, from the Permian down to the Cambrian. Most of it is associated with crude oil, either in solution or in the form of gas-cap accumulations. The Carthage pool is in northeastern Texas. Here 10 different layers in the Trinity division of the Cretaceous system have been found productive. The cumulative total production to the end of 1957 was somewhat over 5×10^{12} ft^3. During 1956 gas production in the whole state of Texas was about 6×10^{12} ft^3. Of this amount two-thirds was gas-well gas and one-third casinghead gas. Four-fifths was used as fuel (either on the lease or in transmission lines) and for the making of carbon black. Cretaceous rocks are the principal reservoir rock in northern Louisiana and in Mississippi The large Monroe dry-gas field in northeastern Louisiana had produced about $5\ 5 \times 10^{12}$ ft^3 up to the close of 1957. Throughout the Rocky Mountain states various layers in the Cretaceous system account for most of the gas. There are many dry-gas pools Outstanding in importance are the Blanco (northwestern New Mexico), the Baxter Basin (southwestern Wyoming), and the Cedar Creek (southeastern Montana) In California gas production is derived from various layers in the Tertiary system. Although about three-fourths of the gas is associated with oil reservoirs, there are a number of dry-gas fields. The largest is the Trico field in San Joaquin basin This field produced 9×10^9 ft^3 during 1957.

Related products and problems. A method has now been perfected to change methane into liquid form. This means that the Middle East with its vast potential can send out tanker loads of liquefied methane to all parts of the world to compete with other forms of mineral fuel Such impending development demands a reappraisal of all natural gas resources.

Helium, the most valuable by-product of natural gas, has been found in some pools. The Rattlesnake pool in New Mexico contains 7.5%, the highest found by 1958 Reserves in New Mexico total about 1×10^9 ft^3, those in Oklahoma about 10×10^9, and those in Texas about 2×10^9. *See* MINERAL FUEL AREAS; PETROLEUM; PETROLEUM GEOLOGY. [W.A.V.W.]

Bibliography: R. L. Huntington, *Natural Gas and Natural Gasoline*, McGraw-Hill, 1950; *Fuel*, U.S. Bur. Mines Minerals Yearbook 1957, vol. 2, 1959.

Nautical Almanac

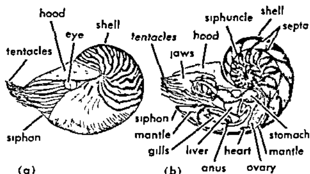
A book published annually by the governments of the principal maritime nations, in the United States beginning in 1855, containing the astronomical data required for navigation by observations of celestial objects. With the aid of the volume, a sextant, and knowledge of Greenwich mean time, a navigator can find his latitude and longitude, fixing his position within 1-2 miles (*see* NAVIGATION).

The modern Nautical Almanac tabulates Greenwich hour angle and declination of the Sun, Moon, Venus, Mars, Jupiter, and Saturn, for every hour of the year, and the sidereal hour angle and declination of about 60 stars for every third day (*see* ASTRONOMICAL COORDINATE SYSTEMS). Also given are times of sunrise, sunset, moonrise, moonset, and twilight, with other astronomical phenomena.

A similar publication is especially designed to facilitate air navigation. *See* AIR ALMANAC; *see also* EPHEMERIS. [C M C.]

Nautiloidea

An order of the tetrabranchiate cephalopods considered to be the most primitive of the class. The shells are straight or coiled and chambered with curved transverse septa; they have simple sutures and simple exterior sculpture. There are about 300 genera and more than 2500 species. The first records are from the Upper Cambrian. The group reached its zenith in the Ordovician and had largely disappeared by the Triassic. The sole living genus, *Nautilus* (see illustration), is represented in mod-



The pearly nautilus, *Nautilus pompilius* (a) External features. Shell to 10 in. in diameter (b) Internal structure; shell and mantle (except at siphon) cut away to midline; jaws, tongue, and radula shown in median section, two left gills removed (From T I Storer and R L. Usinger, General Zoology, 3d ed, McGraw-Hill, 1957)

ern seas by six species all occurring in tropical regions of the Indo-Pacific where they live in depths of around 500 ft It has four gills, four kidneys, numerous suckerless retractile tentacles and a funnel formed of two contiguous flaps. The shell is external, smooth and coiled, and is sold commercially. *See* CEPHALOPODA; TETRABRANCHIA. [C.L.V.]

Nautilus

The sole surviving genus of the family Nautilidae, class Cephalopoda, phylum Mollusca. There are about 2500 fossil species dating from the Cambrian geologic period to the Recent. Among living forms only one species is well known, *Nautilus pompilius*, the chambered or pearl nautilus. However, various authorities recognize from one to five additional living species.

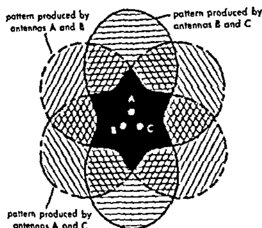


Fig 1 Navaglobe antenna location and field pattern

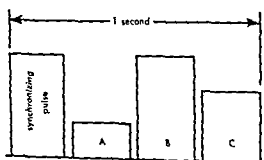


Fig 2 Navaglobe keying cycle.

from the station Ambiguity (multiples of 360° phase difference) is resolved by measuring the phase of a 100-cycle signal generated through the use of this same oscillator and a 100-cycle modulation from the ground station

The Navaglobe-Navarho receiver is of special design It incorporates (in highly refined form) the principles of a vector ratiometer with three current carrying coils mounted at a 120° space relation to each other Currents proportional to the received amplitudes of the A, B and C signals are applied in sequence to each of the three coils A magnetic needle placed at the center of these coils would assume the direction of the vector resultant of the three magnetic fields and hence, would indicate the direction of the transmitting station

Navarho-Navaglobe was designed to operate on a frequency of 90–110 *kc* and has an accuracy of about ½° in bearing Distance measurement accuracy is a function of the time which has elapsed from the time of setting of the phase of the oscillator Tests have shown that an accuracy of plus or minus 3 miles may be achieved *See* NAVIGATION SYSTEMS, ELECTRONIC.

[P C S.]

Navier-Stokes equations

Three scalar partial differential equations that describe conservation of momentum for the motion of a viscous, incompressible fluid They may be expressed vectorially as one equation

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p + \rho \mathbf{f} + \mu \nabla^2 \mathbf{v}$$

where ρ is fluid density, $\mathbf{v}$ is fluid velocity vector, p is fluid pressure, $\mathbf{f}$ is body force (such as gravity) per unit mass, μ is fluid viscosity coefficient, and t is time (*see* NEWTONIAN FLUID). These equations, together with the continuity relation, $\nabla \cdot \mathbf{v} = 0$, and suitable boundary conditions determine the flow field; for example, $\mathbf{v}$ and p are determined as functions of position in space and of time. One of these boundary conditions is that of no slip at the surface of a body, that is, the fluid immediately at the body surface "sticks" to it and thus has the same velocity as the surface itself

Few mathematical solutions are known to this complicated set of nonlinear partial differential equations except for simple geometries. The importance of viscosity in determining the flow depends on the relative size of the body (*see* REYNOLDS NUMBER). Approximations to the Navier-Stokes equations for small Reynolds number *Re* give good results For $Re < 1$, the acceleration forces, those on the left-hand side of the equation, are negligible, leaving only linear terms on the right Such an approximation is called a Stokes-

mation as well as even further approximations. For $Re \gg 1$, the effects of viscosity are confined to a thin layer near the surface of bodies in the fluid

use at all (*see* D'ALEMBERT'S PARADOX).

For a compressible, viscous fluid, the viscous term $\mu \nabla^2 \mathbf{v}$ must be replaced by the divergence of the viscous stress tensor in which the bulk viscosity coefficient λ occurs Using rectangular cartesian coordinates x_1, x_2, x_3, the force in the x_i direction ($i = 1,2,3$) can be written as

$$\sum_{j=1,2,3} \frac{\partial \tau_{ij}}{\partial x_j}$$

where $\tau_{ij} = (\lambda - \tfrac{2}{3}\mu)\,\delta_{ij}(\nabla \cdot \mathbf{v}) + \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$

and $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Here μ and λ are functions of temperature and, in liquids, of pressure. *See* FLUID-FLOW PRINCIPLES.

[A.E.BR]

Navigation

The process of directing the movement of a craft from one place to another The craft may be a ship, aircraft, missile, spacecraft, land vehicle, or in the broadest sense, any object requiring direction or capable of being directed

hurricanes and tropical storms, and systems for routing ships that have resulted in quicker ocean travel.

Wave forecasting. Meteorologists now regularly study wind data and the marine field (fetch) over which the wind blows to forecast the height, period, wave length, and decay period of ocean waves. This information is widely used by the U.S. Navy for its importance in amphibious and supply operations.

Two situations from World War II will illustrate. The beaches of Normandy have slopes with about 1 ft of rise for 100 ft or more of beach. A 6-ft wave breaking in 8 ft of water would dissipate its energy over 800 ft of beach. By contrast, Iwo Jima's beach slopes are very steep, about 1 ft of rise for 10 ft of beach. A wave breaking at 8 ft has only 80 ft of beach area in which to expend its energy. Thus, landing conditions are correspondingly rougher on Iwo Jima under various wave conditions. Forecasting of this sort is a development of naval meteorology.

Radar and sonar. Naval meteorology has been aiding the use of radar at sea since research found that choppy surface conditions created radar interference, called sea clutter. Wind and other data are now used to forecast the quality of marine radar observation at a given time.

The range of sonar detection depends, among other things, on the thermal patterns of the sea's surface layer (*see* SONAR). From wind and wave data, the depth and character of the thermal layer can be estimated and the range of the sonar sounding predicted.

Hurricane forecasting. The largest single expansion in naval meteorology since World War II has been concerned with hurricanes and tropical storms The U.S Navy, Air Force, and Weather Bureau have joined to study, track, and forecast these phenomena. Manned aircraft penetrate the center of hurricanes and collect data that, with radar observations, have provided much new information about their structure.

Another technique is used by the U.S. Navy's Bureau of Aeronautics. Instrumented buoys are placed so as to give a detailed picture of temperature, pressure, humidity, and wind immediately around a hurricane. The Navy also developed a technique for dropping large radar-reflective plastic spheres into hurricanes, where they move on the sea surface with the eye of the storm The spheres are visible with radar for 200 miles.

Transosonde. An important data collection method called transosonde is providing information, often previously unavailable, about the atmosphere over oceans. Huge plastic balloons carrying instrument packages are released to float over ocean basins at 30,000 ft. In 1957 the U.S. Navy inaugurated daily balloon launchings from Japan and the east coast of the United States. The prevailing westerly winds commonly carry the balloons over the Pacific and the Atlantic oceans in two or three days.

Routing ships. The U.S. Naval Hydrographic Office offers a regular forecasting service to help ocean-going ships find the best route offered by current weather conditions. Ship captains receive new information every 24 hours, and thousands of dollars in shipping costs are saved every year. The forecasters, with much aid from machine computers, consider the combined effect of wind, weather, wave height, and ship characteristics in calculating the quickest route. *See* AEROLOGY; AERONAUTICAL METEOROLOGY. [E G.D]

Naval stores

Rosin and turpentine obtained from the industrial processing of resinous (pine) woods. See ROSIN; TURPENTINE. Southern slash pine is the source of much of the domestic naval stores. Rosin and turpentine are obtained either from the exudate from incisions in the trees or from the destructive or steam distillation of wood. Products from the former process are called gum rosin and gum turpentine, and from the latter, wood rosin and wood turpentine.

In the distillation method, the wood is either steam distilled or, more often, destructively distilled *See* WOOD CHEMICALS.

In the extraction method, comminuted wood is first steamed under pressure to remove turpentine or is treated directly with a lower boiling solvent which can be fractionated from the turpentine. The rosin is usually separated by treatment with caustic and subsequently liberated by the addition of dilute sulfuric acid

For gum naval stores, the gum exudate from incisions made in live pine trees is collected and is steam distilled for turpentine The residue consists mainly of rosin which is purified in much the same way as that obtained by other processes

[E L.S.]

Navarho

A long-distance continuous-wave navigation system providing simultaneous bearing and distance information. Navarho incorporates Navaglobe, which is the portion of the system providing bearing.

Navaglobe utilizes three antennas located at the apex of a triangle as in Fig 1. The spacing between antennas is approximately 0.4 wavelength. The antennas are energized in pairs. Each pair of antennas produces a dumbbell pattern Each pair is keyed on for ¼ sec, after which an omnidirectional transmission on a frequency 100 cycles from the signal frequencies is transmitted for synchronization purposes (see Fig 2)

On the aircraft, a highly accurate frequency standard with a stability of at least one part in 10^9 is used to produce a measurement of distance The phase of the oscillator driven by this standard is set at the time of aircraft departure or at the time of reaching a destination which is at a known distance from the ground facility. The phase difference between this oscillator and the omnidirectional signal is therefore proportional to distance

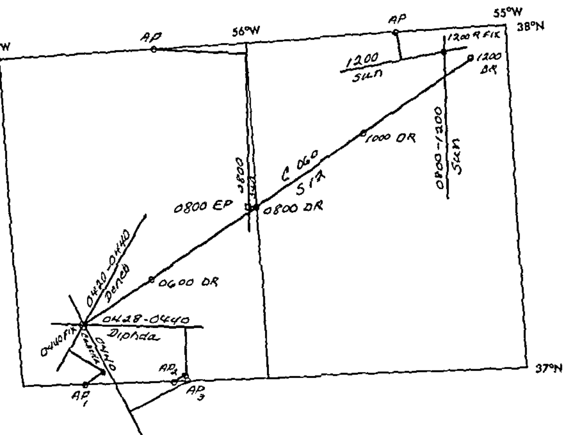


Fig 1. A typical navigational plot at sea

The direction in which a craft is pointed, by whatever reference direction it is indicated, is called heading. The intended direction of motion is called the course. The actual direction from a point of departure to a point of arrival is called the course made good. The actual path followed is the track. This term is also widely used by air navigators to refer to either an instantaneous or the average direction of the track. A bearing is the direction of one terrestrial point from another. It generally refers to the direction of an object as viewed from a craft. The bearing of a celestial object is called its azimuth. This term is occasionally used by navigators to refer to directions on the earth, but it is usually considered better practice in navigation to restrict the use of azimuth to the directions of celestial bodies and use bearing for directions of terrestrial points.

Horizontal directions are customarily stated in whole degrees, although half and quarter degrees may be used, and tenths of a degree are commonly used by marine navigators for indicating azimuths. Minutes of arc are seldom used for expressing directions in navigation and seconds of arc are almost never used.

Reference directions are determined by means of a compass. For many centuries, the primary directional instrument has been a magnetic compass which depends for its operation on the attraction of the earth's magnetic field for a magnetic element mounted so as to be free to turn in any horizontal direction. More recently, other compasses have been developed, notably the north-seeking gyrocompass. See COMPASS, MAGNETIC; DEAD RECKONING, GYROCOMPASS

Distance. This is usually expressed by the navigator in nautical miles, to a precision of integral miles by the air navigator, and tenths of a mile by other navigators. The international nautical mile, now standard in nearly all maritime nations, is 1852 meters by definition. This is 6076.115486 ft, as adopted by the United States on July 1, 1954, but using the recently adopted relationship 1 in. = 2.54 cm. The value approximates the length of one minute of arc of a great circle on the surface of the earth, and so is particularly convenient in navigation because the latitude scale can be used as a mile scale without introducing a large error. Shorter distances are generally indicated in yards, or in meters by countries using the metric system.

The principal maritime nation not accepting the international nautical mile is Great Britain, which uses a value of 6080 ft for instrument calibration.

Two other units are occasionally used for indicating distance. One is the cable, which, in the United States Navy, is 720 ft. In Great Britain, where the unit is in more common use, it is 608 ft, or exactly one-tenth of a nautical mile. The other

An adjective is often used with the word navigation to indicate type of craft or the element in which it moves Thus, examples are: air navigation for aircraft, space navigation for spacecraft, marine navigation for watercraft, and land navigation for land vehicles. Other adjectives may refer to location, as in the case of river navigation, ocean navigation, or polar navigation for polar areas. Still others refer to the primary method being used, as radar navigation or inertial navigation. *See* AIR NAVIGATION; CELESTIAL NAVIGATION; MARINE NAVIGATION; POLAR NAVIGATION. *See also* NAVIGATION SYSTEMS, ELECTRONIC.

Successful navigation involves both art and science. As instruments and other navigational aids have become more complicated, an increasing proportion of the development has been shifted from the practicing navigator to the navigational scientist who aids in drawing together the applications of principles from such sciences as astronomy, cartography, electronics, geodesy, magnetism, mathematics, meteorology, oceanography, physics, and surveying. Such applications aim to aid in explaining navigational phenomena and in developing improvements in speed, accuracy, or lessened drudgery in practicing the art of navigation

BASES AND DEFINITIONS

Navigation as related to the direction of the movements of a craft usually involves determination of position, direction, and distance. When the element of time is included with distance, rate of motion, or speed, becomes available.

Position. Several kinds of position are of interest to the navigator Primarily, he is interested in the position of his craft. A reliable position determined from information external to his craft, as by observation of landmarks or celestial objects, is called a fix. This is generally determined by one or more simultaneous, or nearly simultaneous, nonparallel lines of position, the common intersection of the lines being considered the fix The individual lines of position might be arcs of great circles, small circles, hyperbolas, or other figures If nonsimultaneous lines of position are available, they can be adjusted for the craft's motion between observations to provide a running fix of somewhat reduced reliability, because of the possible error in the estimation of the craft's motion between observations.

A position based upon incomplete information, or data of questionable reliability, may be termed an estimated position However, a position determined by advancing a reliable former position is generally called a dead reckoning position. Many marine navigators limit the use of this expression to positions determined by using speed through the water (without allowance for wind effect) along the course steered. "Estimated position," then, is reserved for positions determined by adding to this motion the estimated effect of wind and current. When inconsistent data are available, a common condition in practical navigation, analysis of available data establishes the most probable position

The average offset of a craft by wind or current since the last fix can be determined by comparison of a fix with the position the craft would have occupied had there been no wind or current. It is for this reason, in part, that many marine navigators prefer to make a careful distinction between dead reckoning and estimated positions. In air navigation, the estimated effect of wind is invariably considered in determining a dead reckoning position The no-wind position, sometimes called air position, is used for determining average wind since the last fix.

Other positions frequently of interest to a navigator using celestial bodies for determining a line of position are the assumed position of his craft at the time of observation, and the geographical position of the celestial body at the same time. The former is an arbitrarily selected position in the general vicinity of the actual position, chosen to aid in computing the required positional data. The latter is the terrestrial position at which the observed celestial body is vertically overhead at the instant of observation. A typical navigational plot at sea, showing several kinds of position, is shown in Fig. 1. In this illustration, AP is assumed position, DR is dead reckoning position, EP is estimated position, C is course, and S is speed.

Whatever the nature of the position, it is generally stated in terms of geodetic latitude and geodetic longitude In air navigation, position is customarily stated to a precision of one minute of arc in each coordinate. In marine navigation, the precision is customarily one-tenth of a minute of arc. Position of a craft might also be stated relative to an established position, as "close aboard buoy 2CB," or "1 7 miles bearing 168° from Chesapeake Lightship," or "over the center of Golden Gate Bridge" *See* COORDINATE SYSTEMS, TERRESTRIAL.

Direction. This is customarily expressed as an angular distance from a reference direction. For most purposes of navigation, north is the reference direction, but the forward direction of the longitudinal axis of the craft is used when a relative direction is stated Other directions, such as south or the direction of the prime vertical circle (east or west), occasionally serve as references.

North may be defined in a variety of ways "True" directions are related to geographical north (the north direction along the meridian), magnetic directions to magnetic north, compass directions to north as indicated by a magnetic compass, and gyro directions to north as indicated by a north-seeking gyro compass. Sometimes, particularly in high latitudes, an arbitrary grid of parallel lines is placed on the chart, and directions relative to "grid north" are called grid directions Directions may refer to either a great circle or to a rhumb line, and it is considered good practice to state which is intended where reasonable doubt might exist.

Fig 1 A typical navigational plot at sea

The direction in which a craft is pointed, by whatever reference direction it is indicated, is called heading. The intended direction of motion is called the course. The actual direction from a point of departure to a point of arrival is called the course made good. The actual path followed is the track. This term is also widely used by air navigators to refer to either an instantaneous or the average direction of the track. A bearing is the direction of one terrestrial [illegible]

[illegible] called its azimuth. This term is occasionally used by navigators to refer to directions on the earth, but it is usually considered better practice in navigation to restrict the use of azimuth to the directions of celestial bodies and use bearing for directions of terrestrial points.

Horizontal directions are customarily stated in whole degrees, although half and quarter degrees may be used, and tenths of a degree are commonly used by marine navigators for indicating azimuths. Minutes of arc are seldom used for expressing directions in navigation, and seconds of arc are almost never used.

Reference directions are determined by means of a compass. For man[illegible] rectio[illegible] which [illegible]

the earth's magnetic field for a magnetic element mounted so as to be free to turn in any horizontal direction. More recently, other compasses have been developed, notably the north-seeking gyrocompass. *See* COMPASS, MAGNETIC; DEAD RECKONING; GYROCOMPASS.

Distance. This is usually expressed by the navigator in nautical miles, to a precision of integral miles by the air navigator, and tenths of a mile by other navigators. The international nautical mile, now standard in nearly all maritime nations, is 1852 meters by definition. This is 6076.115486 ft, as adopted by the United States on July 1, 1954, but using the recently adopted relationship 1 in. = 2.54 cm. The value approximates the length of one minute of arc of a great circle on the surface of the earth, and so is particularly convenient in navigation because the latitude scale can be used as a mile scale without introducing a large error. Shorter distances are generally indicated in yards, or in meters by countries using the metric system.

The principal maritime nation not accepting the international nautical mile is Great Britain, which uses a value of 6080 ft for instrument calibration.

Two other units are occasionally used for indicating distance. One is the cable, which, in the United States Navy, is 720 ft. In Great Britain, where the unit is in more common use, it is 608 ft, or exactly one tenth of a nautical mile. The other

An adjective is often used with the word navigation to indicate type of craft or the element in which it moves. Thus, examples are: air navigation for aircraft, space navigation for spacecraft, marine navigation for watercraft, and land navigation for land vehicles. Other adjectives may refer to location, as in the case of river navigation, ocean navigation, or polar navigation for polar areas. Still others refer to the primary method being used, as radar navigation or inertial navigation. *See* AIR NAVIGATION; CELESTIAL NAVIGATION; MARINE NAVIGATION; POLAR NAVIGATION. *See also* NAVIGATION SYSTEMS, ELECTRONIC.

Successful navigation involves both art and science. As instruments and other navigational aids have become more complicated, an increasing proportion of the development has been shifted from the practicing navigator to the navigational scientist who aids in drawing together the applications of principles from such sciences as astronomy, cartography, electronics, geodesy, magnetism, mathematics, meteorology, oceanography, physics, and surveying. Such applications aim to aid in explaining navigational phenomena and in developing improvements in speed, accuracy, or lessened drudgery in practicing the art of navigation.

BASES AND DEFINITIONS

Navigation as related to the direction of the movements of a craft usually involves determination of position, direction, and distance. When the element of time is included with distance, rate of motion, or speed, becomes available.

Position. Several kinds of position are of interest to the navigator. Primarily, he is interested in the position of his craft. A reliable position determined from information external to his craft, as by observation of landmarks or celestial objects, is called a fix. This is generally determined by one or more simultaneous, or nearly simultaneous, nonparallel lines of position, the common intersection of the lines being considered the fix. The individual lines of position might be arcs of great circles, small circles, hyperbolas, or other figures. If nonsimultaneous lines of position are available, they can be adjusted for the craft's motion between observations to provide a running fix of somewhat reduced reliability, because of the possible error in the estimation of the craft's motion between observations.

A position based upon incomplete information, or data of questionable reliability, may be termed an estimated position. However, a position determined by advancing a reliable former position is generally called a dead reckoning position. Many marine navigators limit the use of this expression to positions determined by using speed through the water (without allowance for wind effect) along the course steered. "Estimated position," then, is reserved for positions determined by adding to this motion the estimated effect of wind and current. When inconsistent data are available, a common condition in practical navigation, analysis of available data establishes the most probable position.

The average offset of a craft by wind or current since the last fix can be determined by comparison of a fix with the position the craft would have occupied had there been no wind or current. It is for this reason, in part, that many marine navigators prefer to make a careful distinction between dead reckoning and estimated positions. In air navigation, the estimated effect of wind is invariably considered in determining a dead reckoning position. The no-wind position, sometimes called air position, is used for determining average wind since the last fix.

Other positions frequently of interest to a navigator using celestial bodies for determining a line of position are the assumed position of his craft at the time of observation, and the geographical position of the celestial body at the same time. The former is an arbitrarily selected position in the general vicinity of the actual position, chosen to aid in computing the required positional data. The latter is the terrestrial position at which the observed celestial body is vertically overhead at the instant of observation. A typical navigational plot at sea, showing several kinds of position, is shown in Fig. 1. In this illustration, AP is assumed position, DR is dead reckoning position, EP is estimated position, C is course, and S is speed.

Whatever the nature of the position, it is generally stated in terms of geodetic latitude and geodetic longitude. In air navigation, position is customarily stated to a precision of one minute of arc in each coordinate In marine navigation, the precision is customarily one-tenth of a minute of arc. Position of a craft might also be stated relative to an established position, as "close aboard buoy 2CB," or "1 7 miles bearing 168° from Chesapeake Lightship," or "over the center of Golden Gate Bridge." *See* COORDINATE SYSTEMS, TERRESTRIAL.

Direction. This is customarily expressed as an angular distance from a reference direction. For most purposes of navigation, north is the reference direction, but the forward direction of the longitudinal axis of the craft is used when a relative direction is stated Other directions, such as south or the direction of the prime vertical circle (east or west), occasionally serve as references

North may be defined in a variety of ways "True" directions are related to geographical north (the north direction along the meridian), magnetic directions to magnetic north, compass directions to north as indicated by a magnetic compass, and gyro directions to north as indicated by a north-seeking gyro compass Sometimes, particularly in high latitudes, an arbitrary grid of parallel lines is placed on the chart, and directions relative to "grid north" are called grid directions. Directions may refer to either a great circle or to a rhumb line, and it is considered good practice to state which is intended where reasonable doubt might exist.

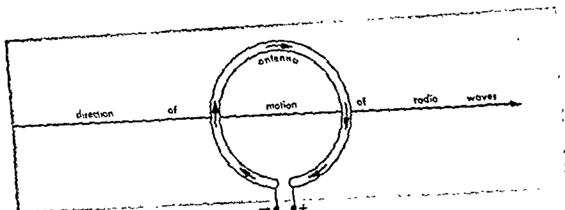


Fig 2 Orientation of a loop antenna for maximum signal strength at the receiver.

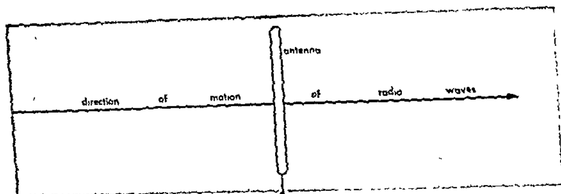


Fig 3 Orientation of a loop antenna for minimum signal strength at the receiver

down. The daily winding of the chronometers is still reported to the commanding officer of U.S Navy vessels *See* TIME.

Navigational electronic uses. Numerous electronic devices or techniques are in various stages of development and use for navigation Navigational aspects of four kinds of such applications are here considered radio piloting, sonar distance and sounding methods, electronic dead reckoning, and techniques for celestial navigation

Radio piloting Various properties of electromagnetic wave propagation and reception may be used for navigation Of these, three kinds have been successfully applied to navigational problems. (1) directional properties of loop antennas, (2) pulse modulation, and (3) rotating directional transmission

1 Directional properties of loop antennas were the first of these properties successfully used. As shown in Fig 2, a wave approaching a loop antenna so as to encounter the sides of the loop successively will induce a current in the loop If the antenna is rotated 90°, so that both sides of the loop are encountered at the same time, as shown in Fig. 3, the effects on the two sides cancel each other, and no current is induced in the antenna. Thus, when the loop antenna is properly mounted and provided with a suitable index and scale, the relative strength of the signal on different orientations provides an indication of the position of the loop with respect to the signal. In the radio direction finder, the scale is graduated so that an accurate reading may be obtained Best results are attained when a minimum-strength signal, called a null, is observed By this means, the direction of the incoming signal can be determined

In the simple installation described, a 180° ambiguity is possible, as signals from opposite directions produce the same effect. The signal strength in various directions is shown in Fig 4 By use of a vertical sense antenna with the loop, this ambiguity can be resolved, as shown by the signal strength diagram of Fig 5. *See* DIRECTION-FINDING EQUIPMENT.

Since the development of the first radio direction finder, many refinements and some innovations

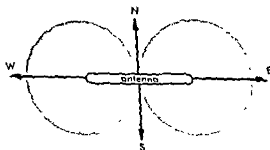


Fig 4 Signal strength of a loop antenna without a sense antenna

unit is the league, now considered chiefly a poetic term This unit is of somewhat indefinite length varying from 2.4 to 4 6 miles.

Depth is usually measured in fathoms in deep water, and in feet in shallow water. One fathom is equal to six feet. Height is generally expressed in feet. Countries using the metric system generally express both depth and height in meters.

Speed. For most purposes of navigation, speed is measured in knots, one knot being equal to one nautical mile per hour. The name of this unit has descended from the chip log, payed out on a knotted line, as used many years ago for determining speed. In recent years, another unit of speed has come into limited use for high-speed aircraft This is the Mach number, after Ernst Mach of Austria. It is the ratio of the speed of a craft to the speed of sound in the medium in which the craft is moving. Thus, it varies with the speed of sound, generally becoming smaller at higher altitudes.

METHODS OF DETERMINING POSITION

All methods of determining position of a craft may be broadly classified under three headings: piloting (or pilotage), dead reckoning, and celestial navigation. The expression electronic navigation is sometimes used in a manner implying a fourth broad classification, but electronics actually only provides another form of energy to supplement visible light and sound in providing data for use in piloting, dead reckoning, and celestial navigation, although the definitions of these divisions, particularly piloting, have had to be somewhat revised The advent of space navigation has further emphasized the thin distinction between piloting and celestial navigation, but the basic difference still exists. Charts and other aids to all navigation are considered at the end of this section.

Piloting. Historically, piloting (so-called by mariners, but called pilotage by aviators) is the earliest form of navigation. It involves the determination of position by and relative to objects external to the craft. In earlier times, these were limited to visible landmarks and prominent underwater features. A vessel was considered to be on soundings as long as it maintained contact with the bottom. When the water became too deep, usually at about the 100-fathom line, the ship was considered to be off soundings. As a vessel proceeded to sea, a last good fix was obtained relative to visible landmarks, and the vessel was said to take departure It then depended entirely upon dead reckoning and celestial navigation until landfall, when land was again sighted, and piloting became possible

More recently, the number of identifiable landmarks has been increased by the addition of man-made objects such as fixed lighthouses and beacons and floating lightships and buoys, as well as by various cultural features. Man-made features intended primarily to assist the navigator in determining the position of his craft or a safe course, or to warn of dangers or obstructions, are called aids to navigation. In contrast, the expression navigational aids refers to all objects, methods, and the like that assist in the navigation of a craft. These include aids to navigation, charts, books, instruments, and others. Artificial earth satellites might be used in this way, and in traveling great distances from the earth, space navigators might use more distant celestial bodies in a similar manner. By means of electronics, such as sonic sounding, piloting techniques have already been extended far from the shores to which they had been confined over the centuries. Electronics also makes accurate time available and provides means for obtaining valuable information relative to storms and other dangers at sea. *See* PILOTING.

Dead reckoning. This method of determining position involves advancing a reliable position the distance a craft is believed to have moved, or will have moved at some future time. Its accuracy depends upon (1) the accuracy of the position being advanced, and (2) the accuracy with which direction and distance are determined. Many marine navigators prefer to use the best determination of course and speed without any allowance for estimated effect of wind or motion of the water.

Because of the uncertainty of motion, with or without allowance for wind and current, the probable error of a dead reckoning position generally increases with time. It may also increase with distance, depending upon the source of error. With the determination of a new position by independent means, a new dead reckoning is usually started. The discrepancy between a dead reckoning position and a fix at the same time is usually attributed to "current" in marine navigation, and to errors in the dead reckoning in other forms of navigation.

Celestial navigation. Unlike piloting, in which position is determined with respect to various objects, celestial navigation involves the use of celestial bodies with respect to the geographical positions they are vertically above. Measurements are made relative to terrestrial references: always the horizontal, and sometimes also a horizontal reference direction, such as north.

Rotational motion of the earth is taken into account by the use of time It is because of the critical effect of time on celestial observations that accurate time is essential when this form of navigation is used Before it was available, navigation at sea was so uncertain that estimates of the time of sighting land, after a long east-west voyage, were often in error by days, and occasionally by weeks Latitude could be determined more accurately Accurate time at sea became available with the invention of the marine chronometer, a high-grade timepiece with a reasonably constant rate under the conditions encountered at sea This was hailed as one of the greatest advances in the long history of navigation. At least three of these instruments are generally carried aboard ship, and it is considered a serious offense if they are permitted to run



Fig 8 An airborne plan position indicator.

This permits return of a stronger signal and coding for more positive identification of the target. This principle is used to provide radar beacons for navigation It is also used with two transponder beacons to provide a system called shoran (short range navigation) developed for use in blind bombing during World War II. An even more accurate system, called hiran (high precision shoran), is now used principally for surveying

Pulse modulation has been used in other systems generally of longer range than radar. In a hyperbolic system such as the American loran (long range navigation) or the British Gee, synchronized pulses are transmitted from two stations, the master station controlling the transmissions of the slave station. The difference in time of reception of the two signals at a craft some distance away is an indication of the difference in distance from the two stations. The locus of all points having the same reading is a hyperbola The ambiguity as to which of the two similar parts of a hyperbola passes through the craft is resolved, one method being by a system of transmission delays that permits identification of the otherwise identical signals from the two stations If a second pair of stations is available (one station of which might be common to both pairs), a fix can be obtained, as shown in Fig. 9. The difference in phase of signals received from different transmitters (providing hy-

[illegible]

ican Raydist system, but the latter is used in the Decca system invented in the United States and developed in Great Britain

A number of variations of the relatively simple [illegible]

Omega [illegible]

3 Rot[illegible]

another [illegible]

done in a manner permitting determination of direction by means of an ordinary radio receiver, so that the special receiving and indicating equipment of the systems described earlier is not needed.

The first such system physically rotated a directional antenna at an established rate, with a distinctive signal being transmitted in all directions at the moment the directional signal, such as a null, was being transmitted in a reference direction. Later, the rotation was performed electrically, the bearing being determined by counting a series of dots and dashes during the so-called keying cycle. The first such system was the German Sonne, installed during World War II as an improvement over the complicated fixed-pattern Elektra system. A British version is called Consol, and an American version Consolan. The Japanese have a version that has not been given a distinctive name.

These systems are intended for general navigational purposes either in the air or aboard ship A specialized system has been developed to replace the Adcock antenna system of four-course ranges for airways. Two versions, each permitting indication of direction of transmitter by dial, and distance by counter, have been combined to form a system called Vortac (from VOR-DME, the designation of the omnirange-distance-measuring equipment system developed for civilian use, and the Tacan system developed for military use). *See* RADIO RANGE.

Sonar measurements. Sonar equipment uses sonic or ultrasonic signals in water to determine

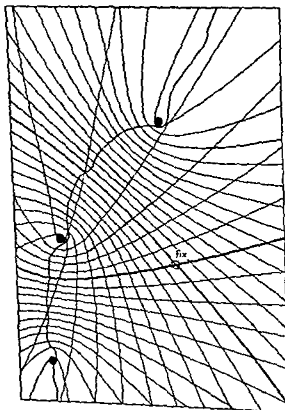


Fig. 9 A fix by a hyperbolic system of navigation

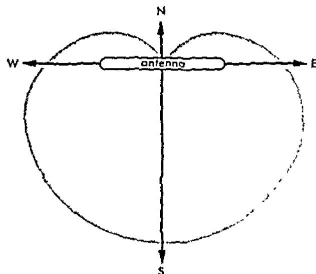


Fig. 5. Signal strength of a loop antenna with a sense antenna.

have taken place, but the basic principle remains the same.

The principle of the loop antenna has been utilized on the ground in antenna systems that radiate signals in a directional pattern. A typical Adcock antenna propagation pattern is shown in Fig. 6. Morse code letters A and N are transmitted in opposite lobes and so timed that when both are received with equal strength, a continuous tone is heard. Such installations, located at various places in the United States, were so oriented that beams of equal strength extended along the airways. By keeping "on the beam," an aviator could stay in the airway. A rotating antenna system (discussed in a later section) has replaced most of these installations. *See* RADIO RANGE

2. When pulse modulation was developed and applied to navigation problems, another long step forward was taken. This technique makes possible

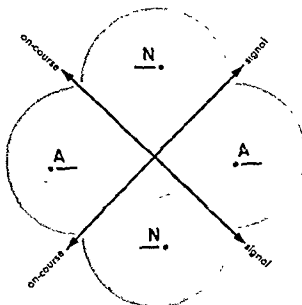


Fig 6. Propagation pattern of an Adcock antenna for air navigation

the transmission of very short bursts of energy separated by comparatively long intervals of silence. In modern pulse systems a typical pulse may last for 1 microsecond (one-millionth of a second) or less, and consecutive pulses may be at intervals of 1000 microseconds or more. This permits relatively high peak power with comparatively low average power.

In radar (radio detection and ranging) pulses are transmitted by a highly directional antenna *which generally rotates several times per minute.* After transmission of a pulse, the transmitter is disconnected from the antenna and a receiver is connected automatically. If the outgoing pulse strikes a suitably reflecting surface, a small amount of energy is returned in the form of an echo, as

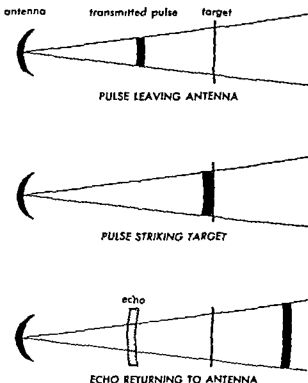


Fig. 7 A radar echo.

shown in Fig. 7 The elapsed time required for the signal to travel to the target and the echo to return is directly proportional to the distance of the target.

Suitable registering devices indicate the distance in miles, yards, or some other linear unit If a circular plan-position indicator is synchronized with the rotating antenna, a maplike indication is obtained, as shown in Fig 8. Thus, both direction *and distance of all targets, both fixed and moving,* are available, as well as a pictorial presentation of the relative positions of all targets within range. The radar equipment may be installed either in the craft or on the ground *See* RADAR

The foregoing system is called primary radar, but with additional functions and components, it becomes secondary. In secondary radar a transponder at the target receives the signal from the craft and automatically transmits a return signal

of polar regions are the transverse Mercator (a Mercator projection with the axis of the cylinder rotated 90°), the modified Lambert conformal (the Lambert conformal projection modified so that there is no gap when the cone is cut along an element and spread out flat) and the polar azimuthal equidistant projection (a nonconformal azimuthal projection with constant scale along any radial

phasis given certain information. Nautical charts are concerned primarily with depths of water and other hydrographic information, and aids to marine navigation, while aeronautical charts emphasize heights and other obstructions to flight, and aeronautical information.

A graticule without other information, other than one or more compass roses to assist in measurement of direction, is called a plotting sheet. A plotting sheet with the latitude or longitude lines left to be filled in by the navigator, so the sheet can be used for virtually any latitude, is called a universal plotting sheet. A graticule having limited information, such as outlines of land areas, is called a plotting chart.

Course lines, lines of position, and other data are usually plotted directly on a chart or plotting sheet. In marine navigation, directions are usually plotted by means of a parallel ruler (a device which can be moved parallel to itself) or a drafting machine (a device combining parallel motion with direction indication). Aviators generally use a simple form of plotter consisting of a protractor and attached straightedge. Dividers are widely used for measuring distances. Compasses are used for drawing circles of visibility of navigational lights and for other purposes.

Other navigational aids. Although many navigational solutions are derived graphically, often directly

solutio

of nav

spheri ... introducing a significant error. Logarithms have been widely used for navigational computations, but in recent years their use has steadily declined as an increased number of tabulated solutions has become available.

For American navigators, most of these tabulated solutions have been provided by the U.S. Navy Hydrographic Office. Examples are H.O. 214, *Tables of Computed Altitude and Azimuth*, the most widely used tables for

of several

tials and

Tables for ... Navigation, the standard for aviators of the United States, Canada, and Great Britain.

Numerous other publications are available to assist the navigator. Examples are nautical and air almanacs of astronomical information of use in navigation, published by the U.S. Naval Observatory and similar offices of other countries; tide and tidal current tables published by the Coast and Geodetic Survey of the U.S. Department of Commerce, and abroad; and coast pilots or sailing directions published by many countries to provide guidance in coastal areas and adjacent waters, and similar air pilots for use by aviators. Notice to Mariners and Notice to Aviators are published at frequent intervals by the U.S. Navy Hydrographic Office to inform navigators of changes affecting charts or publications and other timely navigational information. Similar publications are distributed by other countries. Urgent navigational warnings are broadcast by radio. [A.B.M.]

Bibliography: N. Bowditch, *American Practical Navigator*, U.S. Navy Hydrographic Office, H.O. 9, 1958; J. C. Hill II, T. F. Utegaard and G. Riordan ...

... Whitten and E. Schmid, International symposium on electronic distance-measuring techniques, *J. Geophys. Research*, 65(2): 385–528, 1960.

Navigation instruments

Measuring devices used in ships and aircraft to determine geographic position. The accuracy of such devices has improved steadily since the beginnings of travel, with occasional spectacular new inventions suddenly enlarging man's navigable domain. Notable milestones are the magnetic compass, the sextant, the gyrocompass, radio systems, and inertial guidance.

The magnetic compass was the first, and perhaps the most important instrument in the history of navigation. By furnishing, in any weather, a reliable indication of ship's heading it allowed mariners for the first time to venture many days from land, as the Norsemen proved in their epic voyages to North America about 1000 A.D. See COMPASS, MAGNETIC.

The sextant

of effort

accurate

dawn ... weather, mariners could obtain an accurate fix with no reference but the horizon. See SEXTANT.

The ingenious gyrocompass, invented in 1908, uses gravity and earth's rotation to locate the earth's spin axis (true north). It is much more accurate than the magnetic compass, which it has replaced on most ships (see GYROCOMPASS). Several other gyroscopic instruments have been developed especially for navigation, including the directional gyro and the gyro-stabilized platform. See GYROSCOPE.

The development of electronic

tems

Long-ra

by mea

... of the distance

distance of a target in a manner similar to radar in air. Similar equipment is employed to determine depth of water by echo sounder. Sonar may use electronics in its transmitting, receiving, and amplifying components. *See* SONAR.

Electronic dead reckoning. Two types of electronic dead reckoning systems have been developed during recent years. In one of these, two or more beams of pulse-modulated radio energy are directed obliquely downward from an aircraft. Because of the Doppler effect, the frequency of the echo return from the ground differs slightly from that of the transmitted signal. The amount of the frequency shift is proportional to the speed of the aircraft, and is used as a measure of speed. The beams of radio energy are directed somewhat to the right and left of the longitudinal axis of the aircraft, as well as being tilted downward When two beams are rotated until the Doppler shift is the same for both beams, as determined by a beat frequency oscillator, the drift of the aircraft is measured by the angular difference between the axis of the craft and the relative bearing of the antennas. *See* DOPPLER RADAR.

The other dead reckoning electronic system uses extremely accurate gyroscopes to maintain reference planes, and highly sensitive accelerometers to measure acceleration in various directions The component of acceleration in the direction of motion can be integrated to provide an indication of speed A second integration provides an indication of distance. This inertial navigation has been used successfully in guided missiles, manned aircraft, and ships. This was one of the primary methods of navigation used by the USS *Nautilus* in its historic crossing of the North Pole under the ice pack in 1958. Inertial navigation, in a somewhat different form, is one of the more promising systems proposed for use in space navigation *See* INERTIAL GUIDANCE SYSTEM.

Techniques for celestial navigation. Electronics also has entered the field of celestial navigation. In one application, a star follower, with a photoelectric cell as its heart, automatically "locks on" a star at which it is pointed. With auxiliary equipment, it can provide a continuous indication of altitude in an automatic sextant. By the addition of computing and indicating equipment, this system can be expanded in scope to provide automatic celestial navigation giving continuous indication of position, or steering guidance.

Although a star follower can be made sensitive enough to track stars by day, it does not penetrate even a moderate overcast. One possibility of accomplishing all weather celestial navigation is by means of radio astronomy, using signals of radio frequency originating outside the earth's atmosphere. Infrared signals might also be used Another possibility is the use of artificial earth satellites equipped with radio transmitters When all-weather celestial navigation is achieved, essentially uniform accuracy will be available everywhere on the earth without the need for installing and maintaining transmitters, without the vagaries of transmission through long paths within the earth's atmosphere, and without the possibility of transmitter damage by storm or enemy action, or the jamming of signals by an enemy

The navigator's chart. One of the most important aids available to the navigator is his chart. Even with the development of automatic and semiautomatic navigation systems the chart is still valuable for planning and general information. Some electronic systems, such as those of the hyperbolic type, depend heavily upon charts for interpretation of the readings obtained from the electronic equipment.

A chart is a map intended primarily for navigation. The network of latitude and longitude lines, called the graticule, is invariably drawn according to some system (*see* MAP PROJECTIONS). Map projections are often grouped in three classes, according to whether a likeness of the earth's surface is transferred to (1) a plane (azimuthal), (2) a cone or series of cones (conic), or (3) a cylinder (cylindrical). For navigation, one of the most important properties a projection might possess is that of being conformal, or having the angles around any point correctly represented. When this is true, the scale is the same in all directions around the point, although it might not be uniform in different parts of the projection Other properties of interest are the portrayal of rhumb lines (a rhumb line makes the same oblique angle with all meridians), great circles, and variations in scale.

Nearly all nautical charts are on the Mercator projection, a conformal cylindrical projection in which rhumb lines appear as easily plotted straight lines This projection owes its centuries-old popularity among navigators primarily to the fact that the rhumb line represents a constant true course between points. The scale of this projection varies considerably with latitude, but the rate of change is relatively small until high latitudes are reached, where the projection is generally replaced by a conic or azimuthal type of chart.

The Lambert conformal projection, a conic projection with two standard parallels, is most widely used for aeronautical charts, but has come into limited use for nautical charts. On charts of this projection the scale is nearly uniform, and a straight line is a close approximation of a great circle, which is of growing importance with increases in ranges and in the capability of directly following a great circle

Several other projections offer special advantages for charts. The gnomonic projection a geometric azimuthal projection with points on the surface of the earth conceived as projected by radials from the center of the earth to a tangent plane, is used for great circle sailing. It is the only projection in [illegible] straight lines, but [illegible] being nonconfor[illegible] conformal geometric azimuthal projection, is used for polar charts Other projections used for charts

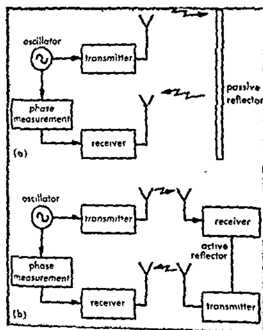


Fig 2 Single path round-trip system (a) With passive reflector. (b) With active reflector

dar, Oboe, Shoran, and the distance-measuring portion of Tacan equipment

Multipath systems There are two types of multiple path systems In one type, emission may be from one or more transmitters, from several an-

[illegible]

ceiver In the second type, radiation may be from a point source and received on several receivers, several antennas associated with one receiver, or many elements of a single antenna associated with one receiver The relative path delay is measured as a phase difference If the paths converge at the vehicle, the navigational parameter is referenced to the earth's geodesy If the paths diverge at the vehicle, the navigational parameter obtained is referred to the heading of the vehicle and must be referenced to the earth's geodesy by use of a magnetic compass or similar device Figure 3 illustrates the basic principle of these multipath systems

The phase difference between the two received signals determines the length of a Since b is a known distance, the angle θ can be found This angle is used to obtain a directional fix on the transmitter

Self-contained systems. There are four types of so-called self-contained navigational systems The term is somewhat misleading, because these systems actually depend on observations made on external phenomena as the basis of their measurements. From these observations, course and speed are determined and used to compute an advanced position from a known position The classification of these systems is based on the phenomena on which the observation is made. Self contained systems include astro devices, which make observations on heavenly bodies; pressure devices, which observe the atmospheric pressure system; Doppler devices, which observe changes in radio frequency due to the Doppler effect; and inertial devices, which measure forces in inertial space

Astro equipment. The astro systems operate by making observations either optically or by radio on celestial bodies and solving the astronomical triangle Star trackers operating through electronic means are expected to assume a prominent place in space navigation. For a discussion of these devices, see NAVIGATION.

Pressure equipment. This equipment, used on all transoceanic flights, makes its observations on atmospheric pressure and includes a barometric altimeter, a radio altimeter, a clock, and a table for use in making computations. In the Northern Hemisphere, winds blow clockwise around high-pressure areas and counterclockwise around low-pressure areas Thus the d[illegible]

is [illegible]

he [illegible]

the radio altimeter A reading of the radio altimeter is made, and after a period of time, during which the barometric altimeter is maintained at a constant altitude, a second reading of the radio altimeter is made. A simple formula is used to determine the vector normal wind. See ALTIMETER, PRESSURE; ALTIMETER, RADIO

Doppler equipment. Radio beams that are transmitted from moving aircraft to the ground and reflected back to the aircraft will have an apparent change of frequency (see DOPPLER EFFECT). The

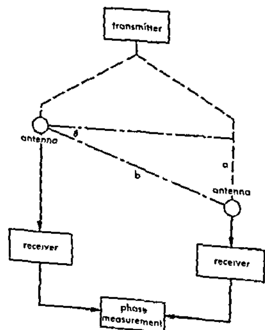


Fig 3 Multipath system Transmission paths (dashed lines) are assumed parallel for long distances

from the vehicle (ship or aircraft) to several different beacons whose locations are known. Navigation near or above land masses is often performed using continuous radar pictures of the terrain itself. *See* NAVIGATION SYSTEMS, ELECTRONIC; RADAR.

Inertial guidance has opened still further the way to long-range navigation by air, sea, and space. *An inertial autonavigator made possible the voyages* of submarines *Nautilus* and *Skate* under the polar ice cap. Inertial systems use precise acceleration sensors to measure vehicle motions. These are tallied to give a continuous reading of vehicle position independent of any external effect. Gyroscopes of extreme accuracy are required for directional reference. In space flight, automatic star-tracking equipment may be used for supervision. *See* INERTIAL GUIDANCE SYSTEM. [R.H.C.]

Navigation systems, electronic

Systems deriving navigationally useful parameters through the application of the electronic sciences. The older term for these systems was "radio aids to navigation." However, several of the newer devices, although employing electronics, do not employ radio waves. Therefore, the term electronic navigation systems is employed to cover all of the devices which use electronics. In electronic navigation systems there is a blending of the sciences of electronics and navigation Therefore, the subject may be discussed from both points of view.

ELECTRONIC CLASSIFICATIONS

From the electronic point of view, it is possible to classify all systems as "classical" or "self-contained" Classical systems employ at least one radio transmitter and one radio receiver. The transmitter emits energy which travels over one or more paths to the receiver The navigational parameter is derived by a measurement made on the delay incurred in the transmission All classical systems are based on the assumption that wave propagation is rectilinear and that the velocity of propagation is constant.

The self-contained systems consist of devices which make observations on certain natural phenomena and computers which derive navigational parameters from the observations.

Classical systems. These have often been classified by the various types of position lines which they produce. but actually they are of only two fundamental types. These are the single-path and *the multiple-path* systems. Two main variations of each of these types exist. The single-path systems measure absolute transmission time and produce circular lines of position The multiple-path systems measure differences in. or otherwise compare. transmission times. Since the locus of all points having a constant difference from two other points is a hyperbola, these systems should produce hyperbolic lines of positions *Actually, many of these* multiple-path systems are so instrumented that they can only determine when the difference in

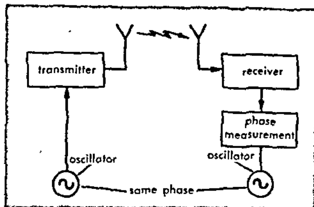


Fig 1. Single-path system.

transmission time is zero (that is, the times are equal). With these systems the line of position is a hyperbola of zero curvature, that is, a straight line. The simplified multiple-path systems are known as radial systems. These radial systems were the first type developed and include direction finders, radio ranges, and the directional portion of radar. The true hyperbolic systems must use transmissions which are diverse in either time or frequency. *See* HYPERBOLIC NAVIGATION SYSTEM.

Single-path systems. The principle of the single-path system is shown in Fig. 1 It employs a transmitter at one point and a receiver at the opposite point *(usually in the vehicle). Transmission occurs* over a single path between the two devices The time required for the radio wave to travel over this path is the important element in determining the navigational parameter. The transmitter frequency, the frequency of the modulation, or the period between successive bursts of carrier energy is accurately maintained Accuracy to better than one part in 10^{10} is desirable. The receiver employs a similar high-stability oscillator, the phase of which is matched to the transmitter's oscillator at a point where the distance between the transmitter and receiver is known accurately Thereafter, the transmitter-receiver oscillator phase difference is a measure of the time taken by the radio wave to traverse the transmitter-to-receiver distance Portable oscillators with the requisite high stability have not been generally available, which accounts for the lack of popularity of the single-path system *See* NAVARHO.

A second type of single-path system employs a round-trip path as shown on Fig 2 The energy from the transmitter is reflected and returned to a receiver located near the transmitter This reflector may be a metallic surface (passive reflector) or an active reflector An active reflector is a receiver with its output connected to a transmitter Upon return, the phase of the reflected energy is compared with that originating at the transmitter The phase difference is a measure of the time, and distance, of the round-trip path The single-path round-trip system is the basis of distance determination in all radars. It is the system employed by Benito, Con-

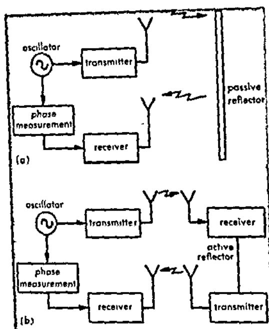


Fig 2 Single-path round-trip system (a) With passive reflector (b) With active reflector

dar, Oboe, Shoran, and the distance measuring portion of Tacan equipment

Multipath systems There are two types of multiple-path systems In one type, emission may be from one or more transmitters, from several antennas associated with one transmitter, or from different sections of a single array associated with one transmitter Reception requires only a single receiver In the second type, radiation may be from a point source and received on several receivers, several antennas associated with one receiver, or many elements of a single antenna associated with one receiver The relative path delay is measured as a phase difference If the paths converge at the vehicle, the navigational parameter is referenced to the earth's geodesy If the paths diverge at the vehicle, the navigational parameter obtained is referred to the heading of the vehicle and must be referenced to the earth's geodesy by use of a magnetic compass or similar device Figure 3 illustrates the basic principle of these multipath systems

The phase difference between the two received signals determines the length of a Since b is a known distance, the angle θ can be found This angle is used to obtain a directional fix on the transmitter

Self-contained systems. There are four types of so-called self contained navigational systems The term is somewhat misleading, because these systems actually depend on observations made on external phenomena as the basis of their measurements From these observations, course and speed are determined and used to compute an advanced position from a known position The classification of these systems is based on the phenomena on which the observation is made. *Self-contained* systems include astro devices, which make observations on heavenly bodies; pressure devices, which observe the atmospheric pressure system; Doppler devices, which observe changes in radio frequency due to the Doppler effect; and inertial devices, which measure forces in inertial space.

Astro equipment. The astro systems operate by making observations either optically or by radio on celestial bodies and solving the astronomical triangle. Star trackers operating through electronic means are expected to assume a prominent place in space navigation. For a discussion of these devices, *see* NAVIGATION.

Pressure equipment. This equipment, used on all transoceanic flights, makes its observations on atmospheric pressure and includes a barometric altimeter, a radio altimeter, a clock, and a table for use in making computations. In the Northern Hemi- [illegible] clockwise around high pressure [illegible] pressure [illegible] an area is an indication of wind direction and intensity. The height of the aircraft above ground is measured by the radio altimeter. A reading of the radio altimeter is made, and after a period of time, during which the barometric altimeter is maintained at a constant altitude, a second reading of the radio altimeter is made A simple formula is used to determine the vector normal wind *See* ALTIMETER, PRESSURE; ALTIMETER, RADIO

Doppler equipment Radio beams that are transmitted from moving aircraft to the ground and reflected back to the aircraft will have an apparent change of frequency (*see* DOPPLER EFFECT). The

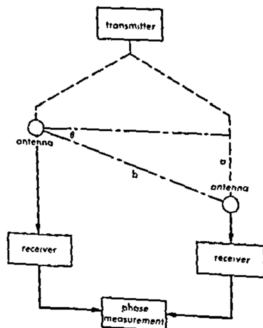


Fig 3 Multipath system Transmission paths (dashed lines) are assumed parallel for long distances

change in frequency is proportional to the ground speed of the aircraft Since maximum Doppler shift is in line with the direction of aircraft movement over the earth's surface, the aircraft drift angle and ground speed are readily determined.

The equipment makes use of a transmitter and receiver operating at centimetric wavelengths. The transmissions may be of the pulse or constant-wave type. A multiple antenna transmits two to four beams. In one design, the antenna array is rotated by a servomechanism; in others the arrays are fixed and information is determined through computation. When the beams are aligned so that the angle between them and the direction of aircraft movement over the ground is equal, the Doppler shifts are equal. The angle between a line bisecting the center of the array and the aircraft heading is the drift angle. The amount of frequency shift is proportional to the true aircraft ground speed One or two rearward-looking beams are employed to compensate for drift in the transmitter frequency.

In order to determine position, a computer takes the determined values of ground speed and drift angle and the aircraft's heading (from a compass or gyro) and computes distance and direction traveled from a previously known position. Therefore, the accuracy of any future position is also a function of the accuracy of the aircraft's directional system. *See* DOPPLER RADAR.

Inertial equipment. Airborne equipment based on Newton's second law can determine aircraft velocity and direction of travel without employment of cooperative ground equipment

From Newton's second law, it is known that the force acting on a body is proportional to its mass and acceleration Acceleration may be determined by measuring the deflection of a spring attached to a known mass. If the force is then doubly integrated, the distance of travel can be determined

The accelerometer must be capable of measuring, without appreciable error, accelerations over a range of 100,000 to 1. Any misalignment of the accelerometer with respect to vertical will cause it to read a component of gravity as the aircraft's acceleration An error of 0 001 in reading acceleration will produce an error of approximately 40 miles in one hour. To attain the required accuracy, the accelerometer is mounted on a platform which is stabilized by gyroscopes having a truly fantastic accuracy

Errors will occur in spite of this equipment, but they are compensated by mechanisms utilizing a principle known as Schuler tuning. In this mechanism, electromechanical systems are given a resonant period of 84 4 minutes The rotational rate of the platform is controlled to be proportional to the integrated acceleration Thus, compensation is achieved by causing the platform to oscillate about the correct vertical position. While errors may mount during one portion of the oscillatory cycle, they will be reversed and not continue to increase with the square of time as they would without Schuler tuning *See* INERTIAL GUIDANCE SYSTEM

Electronic navigation systems

Classical	
Single-path (circular LOP)	
One-way path	
Dectra (distance)	Navarho (distance)
Round-trip path	
Benito	Radar (distance)
Condar	Rebecca
DME (distance-measuring equipment)	Shoran
	Tacan (distance)
Oboe	
Multiple paths (hyperbolic)	
Decca	Omega
Dectra (bearing)	Radux
Gee	Raydist
Loran	
Radial	
Adcock direction finder	Radar bearing
Consol	VOR (vhf omnidirectional radio range)
Four course range	
Glide slope	
Loop direction finder	Wullenberger direction finder
Localizer	
Navaglobe	
Self-contained	
Astro	Inertial
Doppler	Pressure

The table shows the total contribution of electronics to navigation systems.

OPERATIONAL CLASSIFICATIONS

Electronic navigation systems must also be discussed from the standpoint of navigation Navigation is defined as the science or art of conducting vehicles from one point to another. The term is derived from *navis* or ship The term avigation, derived from the Latin *avis*, or bird, should be applied to what is often called air navigation. For years, the mariners have used the term "radio aids to navigation," contending that navigation was an art practiced by the pilot or captain and, therefore, the radio constituted only a device for aiding him. This concept has been passed over by the airmen who use instrument low-approach equipment coupled directly to aircraft controls and other systems for automatic control of the aircraft.

Air operations. Air operations are considered to take place in four zones These are the enroute long-distance zone, the enroute short distance zone, the approach and landing zone, and the airport zone (see Fig 4).

Long-distance enroute zone The equipment employed in this zone has ground facilities located at intervals of not less than 200 miles, such as over oceans or undeveloped areas. It should not be inferred, however, that this equipment is used only where navigational facilities are spaced at intervals of more than 200 miles Many of the same equipments serve useful purposes at much shorter distances. For a discussion of equipment serving in the long-distance enroute zone, *see* AIRBORNE RADAR, DECTRA, DIRECTION-FINDING EQUIPMENT, DOPPLER RADAR; GROUND-POSITION INDICATOR (GPI); INERTIAL GUIDANCE SYSTEM, LORAN, NAVARHO; OMEGA.

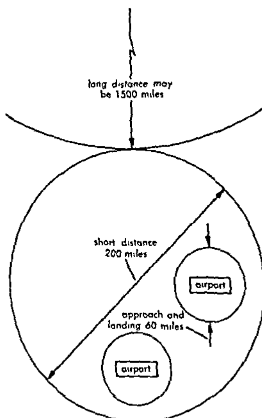


Fig. 4 Air zones of operation

Short distance enroute zone With equipment used in this zone ground facilities may be located at intervals of ...

eve
zon
me
FIC
RANGE, TACAN

Approach and landing zone The equipment used in this zone is primarily in an air space of defined dimensions, together with the runways and water channels used by aircraft arriving at, departing from, and operating in the vicinity of airdromes. See AIRCRAFT LOW-APPROACH SYSTEMS; INSTRUMENT LANDING SYSTEM (ILS), PRECISION APPROACH RADAR (PAR), TERRAIN-CLEARANCE INDICATOR

Airport zone The equipment used in this zone serves aircraft and surface vehicles ...

... of aircraft, when movement within such areas does not constitute a collision hazard with airport traffic See AIRPORT SURFACE ...

... and harbor control See MARINE NAVIGATION

Anti collision. Electronic devices are employed primarily to reduce the risk of collision. This equipment is mainly shipborne radar similar to that employed on aircraft (see AIRBORNE RADAR).

Position fixing. Electronic devices employed for this purpose are classified in three groups: (1) equipment for use at distances over 50 miles, (2) electronic equipment used at distances between 50 and 3 miles, and (3) equipment used for distances under 3 miles.

Equipment used for distances over 50 miles, where an accuracy of 1% is required and 15 minutes is allowable for a position fix, consists of Loran and shipboard direction finders, which take bearings on shore based radio beacons.

Equipment for distances between 50 and 3 miles is employed as an aid upon approaching land, for coastal navigation, and for port approach. It is used where accuracies from one half mile to 200 yards are necessary and from five minutes to one half minute is permissible to obtain a positional fix. Equipment for this purpose consists of shipborne direction finders operating on land based radio beacons, shipboard radars, and radar aids (both active and passive) for special marking of navigational aids, dangers, and shore features Equipment for distances under three miles is employed as an aid to harbor entrance and where 50-yard accuracy and instantaneous position and track fixing are necessary Such equipment consists of high-resolution shipboard radar, radar aids (both active and passive) necessary for special marking of navigational aids dangers, and shore features, and shipboard direction finding employed with radio beacons

Harbor control Harbor control equipment consists largely of high-definition radar installed on shore The services of the harbor radar are transmitted to ships, largely by voice. Some experiments have been made in broadcasting a picture of the radar tube via television The ships employing this service use a standard television receiver to see the location of all ships in the harbor as obtained by the harbor radar [P C S]

Bibliography: J. S Hall (ed), *Radar Aids to Navigation*, 1947, *International Meeting on Marine Aids to Navigation*, 1958; P. C. Sandretto, *Electronic Avigation Engineering*, 1958; P C. Sandretto, *Principle of Electronic Navigation Systems*, VI Convegno Internazionale Delle Communicazioni, Geneva, 1958; War Department, *Air Navigation*, Tech Manual 1-205.

Neanderthal man

A widely distributed form of fossil man taking its name from a skeleton found in 1856 in the Feldhofer cave in the Neander gorge, southeast of Dusseldorf, Germany. This find was recognized as the first example of man differing physically from *Homo sapiens*, modern man In spite of the fact that the face and many other parts were missing, the skeleton revealed a low, broad brain case, continuous arched eyebrow ridges, and projecting occipital region, combined with a brain volume of modern size (Other finds have shown the face to be

change in frequency is proportional to the ground speed of the aircraft. Since maximum Doppler shift is in line with the direction of aircraft movement over the earth's surface, the aircraft drift angle and ground speed are readily determined.

The equipment makes use of a transmitter and receiver operating at centimetric wavelengths. The transmissions may be of the pulse or constant-wave type. A multiple antenna transmits two to four beams. In one design, the antenna array is rotated by a servomechanism; in others the arrays are fixed and information is determined through computation. When the beams are aligned so that the angle between them and the direction of aircraft movement over the ground is equal, the Doppler shifts are equal. The angle between a line bisecting the center of the array and the aircraft heading is the drift angle. The amount of frequency shift is proportional to the true aircraft ground speed. One or two rearward-looking beams are employed to compensate for drift in the transmitter frequency.

In order to determine position, a computer takes the determined values of ground speed and drift angle and the aircraft's heading (from a compass or gyro) and computes distance and direction traveled from a previously known position. Therefore, the accuracy of any future position is also a function of the accuracy of the aircraft's directional system. See DOPPLER RADAR.

Inertial equipment. Airborne equipment based on Newton's second law can determine aircraft velocity and direction of travel without employment of cooperative ground equipment.

From Newton's second law, it is known that the force acting on a body is proportional to its mass and acceleration. Acceleration may be determined by measuring the deflection of a spring attached to a known mass. If the force is then doubly integrated, the distance of travel can be determined.

The accelerometer must be capable of measuring, without appreciable error, accelerations over a range of 100,000 to 1. Any misalignment of the accelerometer with respect to vertical will cause it to read a component of gravity as the aircraft's acceleration. An error of 0.001 in reading acceleration will produce an error of approximately 40 miles in one hour. To attain the required accuracy, the accelerometer is mounted on a platform which is stabilized by gyroscopes having a truly fantastic accuracy.

Errors will occur in spite of this equipment, but they are compensated by mechanisms utilizing a principle known as Schuler tuning. In this mechanism, electromechanical systems are given a resonant period of 84.4 minutes. The rotational rate of the platform is controlled to be proportional to the integrated acceleration. Thus, compensation is achieved by causing the platform to oscillate about the correct vertical position. While errors may mount during one portion of the oscillatory cycle, they will be reversed and not continue to increase with the square of time as they would without Schuler tuning. See INERTIAL GUIDANCE SYSTEM.

Electronic navigation systems

Classical	
Single-path (circular LOP)	
One-way path	
Dectra (distance)	Navarho (distance)
Round-trip path	
Benito	Radar (distance)
Condar	Rebecca
DME (distance-measuring equipment)	Shoran
	Tacan (distance)
Oboe	
Multiple paths (hyperbolic)	
Decca	Omega
Dectra (bearing)	Radux
Gee	Raydist
Loran	
Radial	
Adcock direction finder	Radar bearing
Consol	VOR (vhf omnidirectional radio range)
Four course range	
Glide slope	
Loop direction finder	Wullenberger direction finder
Localizer	
Navaglobe	
Self-contained	
Astro	Inertial
Doppler	Pressure

The table shows the total contribution of electronics to navigation systems.

OPERATIONAL CLASSIFICATIONS

Electronic navigation systems must also be discussed from the standpoint of navigation. Navigation is defined as the science or art of conducting vehicles from one point to another. The term is derived from *navis* or ship. The term avigation, derived from the Latin *avis*, or bird, should be applied to what is often called air navigation. For years, the mariners have used the term "radio aids to navigation," contending that navigation was an art practiced by the pilot or captain and, therefore, the radio constituted only a device for aiding him. This concept has been passed over by the airmen who use instrument low-approach equipment coupled directly to aircraft controls and other systems for automatic control of the aircraft.

Air operations. Air operations are considered to take place in four zones. These are the enroute long-distance zone, the enroute short-distance zone, the approach and landing zone, and the airport zone (see Fig. 4).

Long-distance enroute zone. The equipment employed in this zone has ground facilities located at intervals of not less than 200 miles, such as over oceans or undeveloped areas. It should not be inferred, however, that this equipment is used only where navigational facilities are spaced at intervals of more than 200 miles. Many of the same equipments serve useful purposes at much shorter distances. For a discussion of equipment serving in the long-distance enroute zone, see AIRBORNE RADAR; DECTRA; DIRECTION-FINDING EQUIPMENT; DOPPLER RADAR; GROUND-POSITION INDICATOR (GPI); INERTIAL GUIDANCE SYSTEM; LORAN, NAVARHO; OMEGA.

(a) *Nebalia bipes* (Fabricius), female in lateral aspect, with left valve of shell cut away to show the appendages (after G. O. Sars, *Fauna Norvegiae*, vol 1, 1896). (b) *Nebaliopsis typica* G. O. Sars, in lateral aspect, body distended with large store of food (after H. G. Cannon, *Discovery Rept.* 3, 1931). (c) *Paranebalia longipes* (W. Suhm), fifth thoracic limb (after G. O. Sars, *Challenger Rept* 56, 1887)

on the mud, under or among stones, shells, or weeds *Nebaliella* is a true burrower, with eyes, rostrum, and feelers all adapted for burrowing Any resemblance to the Branchiopoda is superficial The least specialized trunk limbs, those of *Paranebalia*, are essentially similar to the typical biramous malacostracan limbs such as the mysids, although they are shorter and have only faint traces of segmentation The foliaceous limbs of *Nebalia* are highly specialized, an adaptation to its bottom-living and unique filter-feeding habits. The food stream enters anteriorly under the rostral plate, and leaves posteriorly, not the reverse, as stated in textbooks *Nebaliopsis typica*, living at depths [illegible]

of the midgut with reserve food to tide it over periods of scarcity (see illustration b).

The fossil forms differ from living Leptostraca in having, as a rule, three terminal abdominal prongs, the telson being produced as a median style between the furcal rami Some of the fossils are of large size reaching a length of 2 ft. Nothing is known of the structure of their limbs and it is not possible to define their relationship to the living forms *See* LEPTOSTRACA; MALACOSTRACA, PHYLLOCARIDA. [I.C.O.]

Bibliography G. O Sars, *Fauna Norvegiae*, vol [illegible]

Nebula

Originally any diffuse or nebulous astronomical object beyond the limits of our solar system, but now often pertaining only to interstellar clouds of

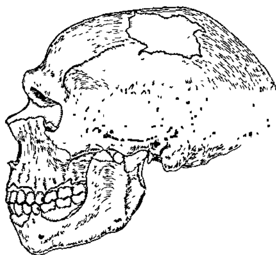


Horsehead Nebula in Orion (IC 434, Barnard 33), photographed in red light with 200-in. telescope, is a cloud of obscuring solid particles between Earth and a gaseous emission nebula. (*Mount Wilson and Palomar Observatories*)

large and extended well forward on the skull, with a large nose, sloping cheekbones, large mouth and dentition, and poorly developed chin.) The skeleton was marked by short limbs and large joints. The specimen was named *Homo neanderthalensis* by W. King This name and classification have been widely accepted. The original specimen is in the *Rheinisches Landesmuseum, in Bonn, Germany*.

Many later discoveries established the validity of the type, showed its wide spread, and allowed its dating. Most may be assigned to the Early Würm glaciation, the first stage of the Fourth Glacial phase. The La Naulette jaw (1866) and the Spy skulls (1886), all from Belgium, demonstrated association with Pleistocene animals In the twentieth century, finds were made in Spain, France (La Chapelle-aux-Saints, Le Moustier, La Quina, and La Ferrassie), Italy, Czechoslovakia, Hungary, western Asia (Shanidar Cave, Iraq; Teshik-Tash cave, Soviet Central Asia), and North Africa (jaw fragments from Tangier, Cyrenaica, and possibly Ethiopia). All these conform well in type to the description above, establishing the nature of the human population of Europe and certain nearby regions during this period. All are associated with some form of Mousterian stone culture.

A smaller number of specimens is of earlier (Third Interglacial) date. These specimens were found in Germany (Ehringsdorf), Italy (Saccopastore near Rome), and Yugoslavia (Krapina in Croatia). They are of basically the same type, in less exaggerated form. The skeleton was apparently less heavily jointed, the skull was less capacious and probably lower but less projecting in the occipital region, and the face was less markedly forward-projecting. Another specimen, possibly still older, is the Montmaurin jaw, from near Toulouse, France; otherwise, any earlier origins of the Neanderthal stock are unknown. Fossil men contemporary with the later Neanderthals (Rhodesian man, Solo man) have by some writers been considered to be Neanderthal variants, or races, but



One type of Neanderthal man, the Old Man of La Chapelle-aux-Saints (From *M. F. Ashley Montagu, An Introduction to Physical Anthropology, 2d ed, 1951*)

the majority regard them as separate developments. *See* RHODESIAN MAN; SOLO MAN.

The relations of the Neanderthals and *Homo sapiens* are likewise problematical. It is evident that the former were replaced in Europe by intruders of modern type, after the Early Würm. Whether the two stocks have a common parentage in the earlier Neanderthals is a debated point. *See* FOSSIL MAN. [W.W.H]

Bibliography: M Boule and H. V. Vallois, *Fossil Men*, 1957; F. C. Howell, The place of Neanderthal man in human evolution, *Am. J. Phys. Anthropol.*, 9(4):379–416, 1951.

Nebaliacea

A small, exclusively marine order of the higher Crustacea (Malacostraca), belonging to the series Leptostraca. Only 5 Recent genera, including less than 20 species and varieties, are known; they are the last surviving members of a group that flourished in Paleozoic times. They are mostly small, 4–12 mm long, bottom-living species that occur in shallow water or at moderate depths; one species is bathypelagic and may be 35–40 mm long

Morphology. *Nebalia* illustrates the diagnostic characters of the order (see illustration *a*). The large bivalve shell or carapace, without any definite hinge line, loosely envelops the thorax and part of the abdomen and, in most cases, quite conceals the thoracic limbs. An adductor muscle in the head region joins the two valves Anteriorly the shell is produced into a movably articulated beak or rostrum. All eight thoracic somites are free and, except in *Nebaliopsis*, are short and crowded, although the suture lines are clearly indicated. The abdomen has seven somites, the last without appendages The telson bears a pair of movably articulated furcal rami

Stalked eyes, antennules, and antennae are well developed The antennal flagellum in the male is as long as the body. The mandible has a large three-segmented palp and a strong molar process; the incisor process is small and simple in *Nebalia*, large and toothed in *Nebaliella* or absent as in *Nebaliopsis* The maxillule has a long obliquely directed palp The eight pairs of thoracic limbs are all similar to each other but differ considerably in structure in the five genera In *Nebalia* the whole limb is much flattened, with a large lamellar epipodite and a lamellar exopodite. In *Paranebalia* (see illustration *c*) the epipodite is minute but both the endopodite and exopodite are long and slender, projecting beyond the edges of the shell The first four pairs of abdominal appendages are biramous and are used in swimming and with the antennules in burrowing; the last two pairs are small, uniramous

In the breeding female, the tip of each thoracic limb, the endopodite, carries a fan of plumose setae bent inward to form a basketlike brood chamber. Development is embryonic, the young hatching at a late stage.

Ecology. Most Nebaliacea live where the bottom deposits are muddy. *Nebalia* lies m[illegible] time

is thermal, like Orion or the Lagoon Nebula, Messier 8, or nonthermal, like the Crab Nebula.

Brightness. Because surface brightness is independent of the distance as long as the eye perceives the object as an extended area, no advantage is gained on objects such as Orion or the Trifid Nebula by using large telescopes unless one wishes to examine small details For small diffuse nebulae and planetaries, a large telescope has considerable advantage For monochromatic radiation, the surface brightness may be expressed in terms of ergs/(sec) (cm²) (unit solid angle), although other units such as magnitudes per square minute of arc have also been used

The brightness of a nebula can also be measured

be measured by photographic photometry, but the most accurate work is done by photoelectric methods, using a spectrum scanner or narrow band pass filters to select monochromatic radiations.

The measurement of the brightness of a nebula is more complicated than that of a star The nebula is an extended surface of nonuniform brightness; hence, the complete description of a nebula in monochromatic radiation would consist of a set of isophotic contours calibrated in terms of intensity units Gaseous and diffuse nebulae show a huge range in surface brightness from objects like Orion to faint wisps barely visible on long exposures with narrow band-pass filters

Distances. If the nebula is associated with a star or a star cluster, its distance may be found by measuring the stellar distances Thus, the distance of the Orion nebula is found by establishing the absolute luminosities of the illuminating stars and the amount of space absorption Then the distance is found from a comparison of the apparent and

... Network Nebula in Cygnus, it is possible to measure the angular rate of expansion and also the velocity in the line of sight, which gives the radial rate of expansion in kilometers per sec The method cannot yet be applied to the planetaries because the rates of expansion are too slow and the objects are too remote For them statistical methods are used in which radial velocities and proper motions are compared together with correlation between angular diameter and distance. Direct determinations are possible for a few objects—the nucleus of NGC 246 which has a dwarf comparison of absolute magnitude M = 7 0 the planetary nebula in the globular cluster M 15, the planetaries near the central bulge of the galaxy, and those in the companion to Andromeda Astrophysical methods ... also been used ...

... plane- ... minute of arc or ... can be observed with a slitless spectrograph (see ASTRONOMICAL SPECTROSCOPY). An image of the nebula is formed in each of its monochromatic radiations. It is found that the radiations of ions of higher excitation such as neon, Ne^{++++}, are always concentrated closer to the central star than are the radiations of ions of lower excitation such as oxygen, O^{+}. The reason is that the higher energy quanta capable of producing highly ionized atoms ...

... lae, ... it is necessary to use a slit spectrograph, preferably one equipped with high-speed cameras.

The spectra of the gaseous nebulae show the recombination lines of hydrogen and helium, but the strongest lines are often some that have never been produced in any terrestrial laboratory. These are the so-called "forbidden" lines of ions of various abundant elements, indicated by brackets. They represent transitions between the metastable levels of the ground configuration. Figure 2 illustrates these transitions for the ions argon [Ar III] and chlorine [Cl III], where numeral III designates a doubly ionized atom, the ionized atoms having ground configurations $3p^3$ and $3p^4$, respectively (see ATOMIC STRUCTURE AND SPECTRA) Transitions of the type ^2P-^2D in p^3 ions or ^1S-^1D in p^2 or p^4 ions are called auroral transitions because this type of forbidden transition is the most important in the Earth's aurora Transitions between the middle metastable term and the ground term give the so-called nebular lines, while jumps from the highest metastable terms to the ground term give trans-auroral lines

Weak recombination lines of oxygen and carbon are observed in a number of planetaries, but the strongest permitted lines (other than hydrogen and helium lines) are certain oxygen [O III] transitions

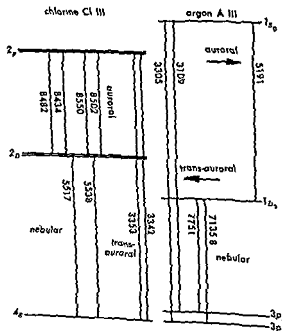


Fig 2 Forbidden-line transitions in doubly ionized chlorine and argon with the energy of each transition given in wave number units.

gas or small particles. Gaseous nebulae may exhibit either an emission or an absorption spectrum. Reflection nebulae and dark obscuring nebulae are composed of small solid particles. Extragalactic nebulae, or external galaxies, are located beyond the boundaries of our own galaxy. *See* ANDROMEDA NEBULA; GALAXY, EXTERNAL; INTERSTELLAR MATTER; NEBULA, GASEOUS; ORION NEBULA. [W.LI.]

Nebula, gaseous

Originally, any fixed, extended, and usually fuzzy, luminous object seen in a telescope. Nebulae are now distinguished from star clouds that can be resolved into individual stars, but earlier workers were unable to differentiate between white nebulae, which are stellar systems so remote as to show no individual stars, and gaseous or diffuse nebulae in our own galaxy.

Extragalactic nebulae are stellar systems comparable with our own galaxy or the Magellanic Clouds in size and number of stars. They are grouped as spirals, ellipticals, or irregulars, and various detailed classification systems have been worked out.

Types of nebulae. The present article deals with gaseous nebulae, clouds of gas like the Network Nebula in Cygnus or dust and gas aggregates such as the Orion Nebula. Gaseous nebulae are members of our galactic system and small compared with its over-all dimensions. Various types of gaseous nebulae have been identified.

Diffuse nebulae. One type of nebula ranges from huge masses of relatively high surface brightness, such as the Orion Nebula, down to faint, milky structures a hundred times less dense that are detectable only with long exposures and special filters. Diffuse nebulae may contain both dust and gas (*see* ORION NEBULA) or may be purely gaseous, like the California Nebula. The nebulosity found in the Pleiades and elsewhere consists of dust with no luminous gas, although it is probable that there is also a great quantity of neutral hydrogen.

Variable nebulae. Variable-brightness nebulae are associated with abnormal variable stars and are frequently fan-shaped in appearance. The best known example is Hubble's Variable Nebula. Another is the nebula associated with T Tauri, which is believed to be a star in the process of formation.

Planetary nebulae. The best-known planetary nebula, although not the brightest, is the Ring Nebula (Fig. 1). Such nebulae often show small greenish disks in the telescope, not unlike the planets Uranus and Neptune, hence the name planetary nebula. They are purely gaseous and are caused to shine by the stars embedded within them.

Extended filamentary radio sources. Examples of radio nebulae are the Network Nebula in Cygnus and NGC 443. Whereas the source of excitation in diffuse, variable, and planetary nebulae is the illuminating star or stars, radio nebulae apparently derive their luminosity from collisions with the surrounding interstellar medium. They emit nonthermal radio-frequency radiation (*see* RADIO ASTRONOMY).

Ex-supernovae. The best-known ex-supernova is the Crab Nebula, which was a supernova in 1054. It is a strong radio source; both the optical and radio-frequency emissions appear to be caused by synchrotron radiation. The explanation of this remarkable object constitutes one of the most challenging problems in astrophysics (*see* SUPERNOVA).

Catalogs. Nebulae are cataloged according to various systems. The first list was due to C. Messier (*see* MESSIER NUMBER). A much more complete list was given by J. L. E. Dreyer in the *New General Catalogue* (abbreviated NGC) and the two Index catalogs. Special lists have been published for diffuse nebulae (S. Cederblad), for new planetaries (R. Minkowski), for faint diffuse nebulae (H. M. Johnson, S. Sharpless), and for different types of radio sources. Many faint extended nebulae are shown in the Palomar charts and in the atlas by G. A. Shajn and B. F. Hase. For the use of the amateur, the Skalnate-Pleso charts are particularly valuable.

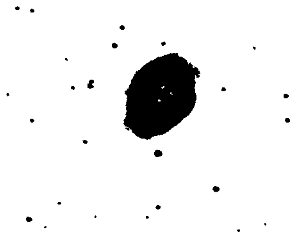


Fig. 1. Ring Nebula NGC 6720. Densely exposed photograph brings out the faint tufts on the outer edge. The space within the ring also shows some structure. The very blue central star is extremely faint. (Photographed with the 60-in. reflector at the Mt. Wilson Observatory, October 26, 1952)

Methods and types of observation. The measurements of positions and sizes are straightforward except for irregular structures that cannot be described in words. Early observations were visual, but by the unaided eye only the brightest nebulae or their most conspicuous features could be detected. Photography has contributed greatly to nebular observation. Most emission nebulae radiate strongly in the red hydrogen line $H\alpha$. Hence by using red-sensitive plates and narrow band-pass filters it is possible to suppress the sky background and register nebulosities of low surface brightness, as was done in Fig. 1. Gaseous nebulae can often be observed with radio telescopes. Thermal radiation is detected at the highest frequencies whereas nonthermal radiation persists to lower frequencies. The measurement of the radio-frequency spectrum is essential for deciding w[illegible] source



Fig 3 Lagoon Nebula NGC 6523 and star cluster NGC 6530 Photographed with a red-sensitive filter and emulsion to isolate the spectral region about the red hydrogen line Hα Note the irregular structure The dark lanes and the globules are caused by solid grains in the neighborhood (*Curtis Schmidt Telescope, University of Michigan*)

... waves appear to occur Also the conventional theory of incompressible turbulence, which gives a definite relation between eddy size and velocity, cannot be applied In some nebulae internal motions are almost certainly complicated by the influence of magnetic fields

Relation to illuminating stars. Except for the nonthermal radio-frequency sources, gaseous nebulae derive their energy from stars near or within them If the star is relatively faint (like T Tauri), the nebula is small, large luminous nebulae are necessarily excited by high-temperature bright stars The hot star ionizes the surrounding gas up to a boundary that is more or less sharp, depending on the density inhomogeneities in the gas Beyond that boundary the gas is neutral and nonluminous, although it may be detectable from its 21-cm radio-frequency emission An example in point is the Lagoon Nebula NGC 6523 which is excited by the star cluster NGC 6530 (Fig 3). The patchy luminous nebula appears to be surrounded by a much larger region of cold, neutral hydrogen The bright O- and B-type stars that excite typical diffuse gaseous nebulae have effective temperatures of 25,000–40,000°K and luminosities of 1,000–10,000 times that of the Sun.

The central stars of planetary nebulae are, by comparison, dwarf stars, although their temperatures range from 25,000°K to perhaps 200,000°K. Their luminosities range from less than that of the Sun to about 200 times that of the Sun; precise figures cannot be given until the distances of the planetaries are accurately established.

Gaseous nebulae in external galaxies. Diffuse nebulae are observed in external galaxies, for example, in the Andromeda Spiral M 31, NGC 6822, and Messier 101; the largest number has been cataloged in the Triangulum Spiral M 33. The largest nebula in this galaxy, NGC 604, has an appearance similar to that of 30 Doradus in the Large Magellanic Cloud, a structure suggestive of the influence of magnetic fields. [L H A.]

Bibliography: L. H. Aller, *Gaseous Nebulae*, 1956, L. H Aller, *Astrophysics: Nuclear Transformations, Stellar Interiors and Nebulae*, 1954; J. Dufay, *Nébuleuses galactiques et matière interstellaire*, 1954; R. Minkowski and O. C. Wilson, *Planetary Nebulae*, 1960

Neck

The communicating column between head and trunk Its bony framework, formed in man by seven cervical vertebrae, is surrounded by heavy muscle groups In front, the esophagus, larynx and trachea, and thyroid lobes occupy the central area, and are covered by thin straplike muscles. On each side of the trachea lie the carotid arteries, the internal jugular veins, and the vagus nerves. The base of the neck contains blood vessels and nerves for supply of the shoulders, arms, and upper trunk,



Human neck (*From J. M Dunlop, Anatomical Diagrams, Macmillan and G Bell, 1935*)

observed in high-excitation planetaries. These [O III] lines are produced by a remarkable fluorescent mechanism discovered by I. S. Bowen. Ions of O^{++} in the $2p^2\ {}^3P_2$ level of the ground configuration absorb the λ 303.78 resonance line of ionized helium and are excited to the $3d^3P_2$ level from which they cascade downward with the emission of observable lines.

The visible Balmer lines of hydrogen and the helium lines are produced by a process of photoionization from the ground level, followed by recombination in one of the highly excited levels with subsequent cascade and the emission of observable lines. For example, the red hydrogen line $H\alpha$ may be produced by the recapture of an electron on the third level with a subsequent jump from the third to the second level. Finally, the atom goes from the second level to the ground level.

On the other hand, the forbidden lines are excited by electron impacts which cause the atoms to rise from the ground term to one of the nearby metastable terms. They return to the ground level with the emission of a forbidden line. The transition probabilities, such as the probability of an atom making the jump in a unit time, are very low. An atom may remain in a metastable level for seconds or even minutes, whereas it will leave an ordinary excited level in a ten-millionth of a second or even less time. The forbidden lines attain such great strength in gaseous nebulae because of their vast extent and because the processes operating to produce the ordinary permitted lines are enormously reduced in efficiency. In a typical planetary nebula the green forbidden lines of oxygen [O III] may be ten times as strong as the hydrogen $H\beta$ line, yet the concentration of hydrogen ions per unit volume may be 10,000 times as great as the concentration of O^{++} ions.

The diffuse galactic nebulae always show relatively low excitation, the Balmer lines are strong, the [O II] lines are often more intense than the [O III] lines, and lines such as [Ar IV] and [Ne V] are absent. In nebulae of low surface brightness, only a few lines can be observed: Hα, and sometimes λ 3727 [O II], and occasionally [O III]. In external galaxies there sometimes exist high-excitation, extended gaseous nebula

Masses, densities, and temperatures. The masses of the nebulae depend on their linear dimensions and their densities Their dimensions are obtained from their angular sizes as soon as their distances are known. Densities are found from the fact that the nebulae are composed mostly of hydrogen. Consider a nebula consisting of a shell of thickness d and outer radius A

The surface brightness in Hα is

$$S_\alpha = 1.61 \times 10^{-19} (N_i N_e / T_e^{3/2}) b_3(T_e) D \exp hRZ^2/n^2kT_e$$

where $D = 3d[1 - (d/A) + (d^2/3A^2)]$

and T_e is the electron temperature, N_i and N_e the ionic and electron densities, respectively, and $b_3(T_e)$ a factor which depends on the electron temperature as follows.

T_e =	5,000	10,000	20,000	40,000
$b_3(T_e)$ =	0.0121	0.116	0.425	0.85

To establish the electron density one needs not only measurements of the surface brightness but also a knowledge of the temperature T_e

One way of estimating T_e is from the energy distribution in the recombination spectrum of hydrogen. The difficulty here is that the temperature is sensitive to the exact distribution of energy, and the effect of the underlying continuum cannot be evaluated easily. One might estimate the electron temperature from the width of the spectral lines, but it would be necessary to separate the effects of the gas kinetic motion from the large-scale mass motion.

The best method involves the use of the relative intensities of the auroral and nebular transitions of a given ion, such as the ${}^1S$-${}^1D$ (λ 4363) and ${}^1D$-${}^3P$ (λ 5007, λ 4959) transitions of [O III]. The relative number of collisional excitations to the ${}^1D_2$ level and to the ${}^1S$ level depends on the velocity distribution of the electrons and hence on the temperature. If the target areas for collisional excitation and the transition probabilities are known, a relation involving temperature, density, and intensity ratio can be found. If the nebular and auroral lines of two ions, both occurring in the same region in the nebula, can be measured, both N_e and T_e can be found independently of the surface brightness measurements. If the nebula has a filamentary structure, the electron density found in this way will be greater than that found from the surface brightness, which represents an average over the space occupied by the nebula

The masses of the planetary nebulae turn out to be about a fifth that of the Sun, although this mass may be somewhat less in some objects. The diffuse nebulae often have masses several times that of the Sun, while the neutral hydrogen clouds in Orion (which are not ionized by the hot stars) have a total mass about 100,000 times that of the Sun

Internal motions. The motions of the gases perpendicular to the line of sight have been found in only a few nebulae, notably the Network Nebula and the Crab Nebula, but motions along the line of sight can be detected by radial velocity measurements with a slit spectrograph and by a Fabry-Perot etalon (*see* INTERFEROMETRY). Use of a multislit consisting of a series of closely placed slits parallel to one another is the most efficient way to observe radial velocity shifts in small nebular regions With this device O C. Wilson has observed many planetaries. The planetaries appear to be expanding, the rate of expansion depending on the degree of ionization of the ion. Thus [Ne V] lines show the smallest expansion rate and [O II] lines show the largest, suggesting that because the degree of ionization depends on the distance from the central star, the material is accelerated on the outer side

dosages of this chemical can be applied safely around the roots of certain established plants.

Water soluble nematocides, such as sodium *N* methyldithiocarbamate (anhydrous), can be applied as soil drenches through irrigation systems. Chemicals, such as *O*-2,4-dichlorophenyl-*O*-diethyl-phosphorothioate, have been applied as emulsions drenched on turf and ornamentals. Solid soil amendments used include calcium cyanamide, urea, and 3,5-dimethyltetrahydro-1,3,5,2*H*-thiadiazine-2-thione. The solids require thorough physical mixing with the soil to be effective

Space fumigants, including methyl bromide, may be used as soil nematocides but are restricted to small scale operations since a gasproof seal or cover is required Methyl bromide is especially useful for sterilizing bags, flats, pots, and other containers or equipment. It has been used to disinfest onion and clover seeds of stem nematode, *Ditylenchus dipsaci*.

Leaf and bud nematodes of the genus *Aphelenchoides* have been controlled by parathion and systox spray applications, but such treatments have very limited usage.

Dip treatments generally employ a combination of hot water and 0 5% formalin solution developed especially for bulb nematode in narcissus and Easter lily bulbs and bulbous iris A slurry of 3-*p*-chlorophenyl 5-methylrhodanine has been used for disinfesting rice seed contaminated with *Aphelenchoides oryzae* See NEMATODA; PESTICIDE

[D J RA]

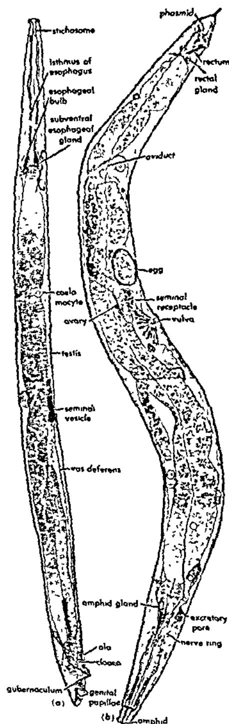
Nematoda

A scientific name commonly applied to a group of unsegmented worms which have been variously recognized as an order, class, and phylum. The legitimacy of the application of both the scientific name and the corresponding English term, nematode has been questioned, because originally the term name was *not applied* in the sense used today, and its derivation from the Greek appears inappropriate [illegible]

pr [illegible]

th [illegible] worms not then known to be parasitic in animals at some stage in their lives. The phylum name Nemates was proposed later for a group including the related free-living [illegible]

[illegible] of which had previously been promoted to both a class and phylum and the latter to a class Subsequently the Nematoda was also promoted to a phylum excluding horsehair worms On the basis of first inclusion and exclusion of groups which are presently placed therein at the rank of a phylum,



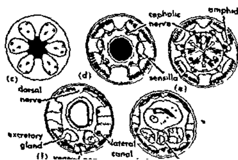


Fig 1 *Pelodera strongyloides* (a) Male (b) Female (c) Cross section through head (d) Stomatal region (e) Esophageal region (f) Postbulbar region (g) Posterior part of male

including the important brachial plexus and subclavian arteries and veins. Subcutaneous tissue, fat, and lymph nodes are plentiful both in deep structures and just beneath the skin. *See* ANATOMY, REGIONAL. [E.C.ST.]

Nectarine

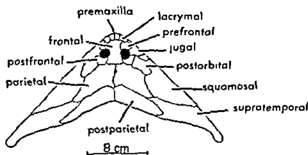
A smooth-fruited, fuzzless form of peach known since the beginning of the Christian era. Fruits usually are somewhat smaller, softer, and richer in flavor than those of the peach. They are highly susceptible to brown rot (*Sclerotinia cinerea*) and surface bruising. Nectarines may originate from seed of the peach, or as a bud sport of the peach. Peaches may, in turn, arise from nectarines. Propagation, adaptation, and cultivation are the same as for the peach. Principal varieties are Quetta, John Rives, Gower, and LaGrande. In North America, cultivation is confined almost exclusively to California Production in 1957 was 36,000 tons, with average returns to growers of $148 a ton *See* FRUIT (BOTANY), FRUIT (TREE); FRUIT (TREE) DISEASES; PEACH. [H B.T.]

Nectonematoidea

An order of parasitic worms belonging to the class Nematomorpha. The order comprises only the genus *Nectonema* which is a pelagic marine organism. *Nectonema* differs from other nematomorph worms by having dorsal and ventral epidermal chords, an open body cavity or pseudocoele, and dorsal and ventral rows of swimming bristles During the parasitic phase of its life cycle, it lives in hermit crabs and true crabs. The larvae are unknown but the adults are occasionally found swimming near the surface of the water. This nematomorph is about 200 mm long. These animals have a cosmopolitan distribution in coastal waters. *See* NEMATOMORPHA. [B.J.MU.]

Nectridia

An order of Carboniferous and Early Permian lepospondylous amphibians characterized by vertebrae in which large fan-shaped hemal arches grow directly downward from the middle of each caudal centrum, the neural arches of the tail, and sometimes of the trunk as well, have a similar shape Most nectridians can be clearly separated into two types. In one, represented by *Urocordylus*, the skull is long and slender, the body and tail elongated and the limbs reduced. In the other type, the head and trunk are broad and flattened and the back corners of the skull table taper into long slender "horns," as seen in *Diplocaulus*. *See* AMPHIBIA FOSSILS; LEPOSPONDYLI. [A.S R.]



Skull of Lower Permian nectridian, *Diplocaulus*. (From E H. Colbert, *Evolution of the Vertebrates*, Wiley, 1955)

Neisseriaceae

A family of bacteria of the order Eubacteriales All known species are parasitic, and several cause disease in man. The bacteria are spherical, gram-negative, and do not form spores The two genera of this family are *Neisseria* and *Veillonella*.

Neisseria. Cells are about 1.0 μ in diameter, and occur commonly in pairs with the adjacent sides flattened Metabolism is strictly aerobic or facultatively anaerobic. *N. gonorrhoeae*, the gonococcus, is an important pathogen of humans, causing the venereal disease gonorrhea. Another pathogen, *N. meningitidis*, the meningococcus, causes epidemic meningitis in man. Other species occur as parasites on the mucous membranes of the respiratory tract of man, but are not known to be pathogenic. *See* GONORRHEA; MENINGITIS; MENINGOCOCCUS.

Veillonella Cells are less than 0 5 μ in diameter and occur in pairs or masses. Metabolism is strictly anaerobic. The best-known species is *V. alcalescens* which carries out a propionic acid fermentation with organic acids as substrates. This organism occurs in large numbers in the saliva of man and animals, where it forms part of the normal flora. It is also found as part of the normal flora in the rumen of sheep. *See* EUBACTERIALES [H C.D.]

Nematocide

A type of chemical used to kill plant-parasitic nematodes Nematocides may be classed as soil fumigants or soil amendments, space fumigants, surface sprays, or dips Soil treatments are commonly used because most plant-pathogenic species spend part or all of their life cycle in the soil, in or about the roots of plants Nematocides may be liquids, gases, or solids, but on a field scale, liquids are most practical. These materials possess a high vapor pressure and volatilize quickly to act as soil fumigants Carbon disulfide, CS_2, and chloropicrin were among the first to be developed for this purpose but are too expensive for general field usage. Also, CS_2 is highly flammable and explosive, and chloropicrin requires a cover or water seal.

The nematocidal properties of 1,3-dichloropropene and 1,2-dibromoethane were discovered between 1943 and 1945, they are the predominant nematocides used at present. These chemicals can be applied to the soil by handgun or machine applicators and require no surface seals or covers beyond cultipacking or rolling the soil after treatment Since most nematocides are phytotoxic, they must be used in the absence of growing plants. The nematocide 1,2-dibromo-3-chloropropane is an exception. Tests since 1954 indicate that nematocidal

dosages of this chemical can be applied safely around the roots of certain established plants.

Water-soluble nematocides, such as sodium [illegible] (anhydrous), can be ap- [illegible] phosphorothioate, [illegible] drenched on turf and ornamentals. Solid soil amendments used include calcium cyanamide, urea, and 3,5-dimethyltetrahydro-1,3,5,2*H*-thiadiazine 2-thione. The solids require thorough physical mixing with the soil to be effective

Space fumigants, including methyl bromide, may be used as soil nematocides but are restricted to small-scale operations since a gasproof seal or cover is required Methyl bromide is especially useful for sterilizing bags, flats, pots, and other containers or equipment. It has been used to disinfest onion and clover seeds of stem nematode, *Ditylenchus dipsaci*

Leaf and bud nematodes of the genus *Aphelenchoides* have been controlled by parathion and systox spray applications but such treatments have very limited usage

Dip treatments generally employ a combination of hot water and 0.5% formalin solution developed especially for bulb nematode in narcissus and Easter lily bulbs and bulbous iris A slurry of 3-*p*-chlorophenyl 5-methylrhodanine has been used for disinfesting rice seed contaminated with *Aphelenchoides oryzae* See NEMATODA, PESTICIDE

[D J RA]

Nematoda

A scientific name commonly applied to a group of unsegmented worms which have been variously recognized as an order, class, and phylum. The legitimacy of the application of both the scientific name and the corresponding English term, nematode has been questioned because originally the term name was not applied in the sense used today, and its derivation from the Greek appears inappropriate according to E Dougherty

The original author called the group Nematoidea, properly deriving the word from the Greek to mean threadlike, but he excluded all worms not then known to be parasitic in animals at some stage in their lives The phylum name Nemates was proposed later for a group including the related free-living and plant parasitic forms and excluding the horsehair worms (Nematomorpha) before the latter were excluded from the Nematoidea or Nematoda the former of which had previously been promoted to both a class and phylum and the latter to a class Subsequently, the Nematoda was also promoted to a phylum excluding horsehair worms On the basis of first inclusion and exclusion of groups which are presently placed therein at the rank of a phylum



Fig 1 *Pelodera strongyloides* (a) Male (b) Female (c) Cross section through head, (d) Stomatal region (e) Esophageal region (f) Postbulbar region (g) Posterior part of male

including the important brachial plexus and subclavian arteries and veins. Subcutaneous tissue, fat, and lymph nodes are plentiful both in deep structures and just beneath the skin. See ANATOMY, REGIONAL. [E.G.ST.]

Nectarine

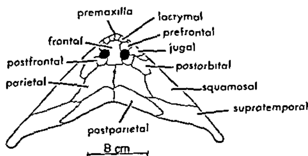
A smooth-fruited, fuzzless form of peach known since the beginning of the Christian era. Fruits usually are somewhat smaller, softer, and richer in flavor than those of the peach. They are highly susceptible to brown rot (*Sclerotinia cinerea*) and surface bruising. Nectarines may originate from seed of the peach, or as a bud sport of the peach. Peaches may, in turn, arise from nectarines. Propagation, adaptation, and cultivation are the same as for the peach. Principal varieties are Quetta, John Rives, Gower, and LaGrande. In North America, cultivation is confined almost exclusively to California. Production in 1957 was 36,000 tons, with average returns to growers of $148 a ton. See FRUIT (BOTANY); FRUIT (TREE); FRUIT (TREE) DISEASES; PEACH. [H.B.T.]

Nectonematoidea

An order of parasitic worms belonging to the class Nematomorpha. The order comprises only the genus *Nectonema* which is a pelagic marine organism. *Nectonema* differs from other nematomorph worms by having dorsal and ventral epidermal chords, an open body cavity or pseudocoele, and dorsal and ventral rows of swimming bristles. During the parasitic phase of its life cycle, it lives in hermit crabs and true crabs. The larvae are unknown but the adults are occasionally found swimming near the surface of the water. This nematomorph is about 200 mm long. These animals have a cosmopolitan distribution in coastal waters. See NEMATOMORPHA. [B.J.MU.]

Nectridia

An order of Carboniferous and Early Permian lepospondylous amphibians characterized by vertebrae in which large fan-shaped hemal arches grow directly downward from the middle of each caudal centrum; the neural arches of the tail, and sometimes of the trunk as well, have a similar shape. Most nectridians can be clearly separated into two types. In one, represented by *Urocordylus*, the skull is long and slender, the body and tail elongated and the limbs reduced. In the other type, the head and trunk are broad and flattened and the back corners of the skull table taper into long slender "horns," as seen in *Diplocaulus*. See AMPHIBIA FOSSILS; LEPOSPONDYLI. [A.S.R.]



Skull of Lower Permian nectridian, *Diplocaulus* (From E. H. Colbert, *Evolution of the Vertebrates*, Wiley, 1955)

Neisseriaceae

A family of bacteria of the order Eubacteriales. All known species are parasitic, and several cause disease in man. The bacteria are spherical, gram-negative, and do not form spores. The two genera of this family are *Neisseria* and *Veillonella*.

Neisseria. Cells are about 1.0 μ in diameter, and occur commonly in pairs with the adjacent sides flattened. Metabolism is strictly aerobic or facultatively anaerobic. *N. gonorrhoeae*, the gonococcus, is an important pathogen of humans, causing the venereal disease gonorrhea. Another pathogen, *N. meningitidis*, the meningococcus, causes epidemic meningitis in man. Other species occur as parasites on the mucous membranes of the respiratory tract of man, but are not known to be pathogenic. See GONORRHEA; MENINGITIS; MENINGOCOCCUS.

Veillonella. Cells are less than 0.5 μ in diameter and occur in pairs or masses. Metabolism is strictly anaerobic. The best-known species is *V. alcalescens* which carries out a propionic acid fermentation with organic acids as substrates. This organism occurs in large numbers in the saliva of man and animals, where it forms part of the normal flora. It is also found as part of the normal flora in the rumen of sheep. See EUBACTERIALES. [H.C.D.]

Nematocide

A type of chemical used to kill plant-parasitic nematodes. Nematocides may be classed as soil fumigants or soil amendments, space fumigants, surface sprays, or dips. Soil treatments are commonly used because most plant-pathogenic species spend part or all of their life cycle in the soil, in or about the roots of plants. Nematocides may be liquids, gases, or solids, but on a field scale, liquids are most practical. These materials possess a high vapor pressure and volatilize quickly to act as soil fumigants. Carbon disulfide, CS_2, and chloropicrin were among the first to be developed for this purpose but are too expensive for general field usage. Also, CS_2 is highly flammable and explosive, and chloropicrin requires a cover or water seal.

The nematocidal properties of 1,3-dichloropropene and 1,2-dibromoethane were discovered between 1943 and 1945, they are the predominant nematocides used at present. These chemicals can be applied to the soil by handgun or machine applicators and require no surface seals or covers beyond cultipacking or rolling the soil after treatment. Since most nematocides are phytotoxic, they

dosages of this chemical can be applied safely around the roots of certain established plants.

Water soluble nematocides, such as sodium N methyldithiocarbamate (anhydrous), can be applied as soil drenches through irrigation systems. Chemicals, such as O-2,4-dichlorophenyl-O-diethyl-phosphorothioate, have been applied as emulsions drenched on turf and ornamentals Solid soil amendments used include calcium cyanamide, urea, and 3,5 dimethyltetrahydro-1,3,5,2H-thiadiazine 2-thione. The solids require thorough physical mixing with the soil to be effective

Space fumigants, including methyl bromide, may be used as soil nematocides but are restricted to small scale operations since a gasproof seal or cover is required Methyl bromide is especially useful for sterilizing bags, flats, pots, and other containers or equipment. It has been used to disinfest onion and clover seeds of stem nematode, *Ditylenchus dipsaci*.

Leaf and bud nematodes of the genus *Aphelenchoides* have been controlled by parathion and systox spray applications, but such treatments have very limited usage.

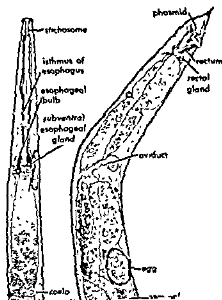
Dip treatments generally employ a combination of hot water and 0 5% formalin solution developed especially for bulb nematode in narcissus and Easter lily bulbs and bulbous iris A slurry of 3-p-chlorophenyl-5-methylrhodanine has been used for disinfesting rice seed contaminated with *Aphelenchoides oryzae* See NEMATODA, PESTICIDE.

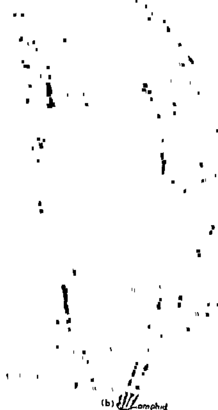
[D J.RA]

Nematoda

A scientific name commonly applied to a group of unsegmented worms which have been variously recognized as an order, class and phylum. The legitimacy of the application of both the scientific name and the corresponding English term, nematode, has been questioned because originally the term name was not applied in the sense used today, and its derivation from the Greek appears inappropriate according to E Dougherty.

The original author called the group Nematoidea, properly deriving the word from the Greek to mean threadlike but he excluded all worms not then known to be parasitic in animals at some stage in their lives The phylum name Nemates was proposed later for a group including the related free-living and plant-parasitic forms and excluding the horsehair worms (Nematomorpha) before the latter were excluded from the Nematoidea or Nematoda, the former of which had previously been promoted to both a class and phylum and the latter to a class Subsequently, the Nematoda was also promoted to a phylum excluding horsehair worms On the basis of first inclusion and exclusion of groups which are presently placed therein at the rank of a phylum,





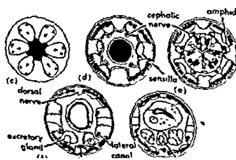


Fig 1 *Pelodera strongyloides* (a) Male, (b) Female (c) Cross section through head, (d) Stomatal region (e) Esophageal region (f) Postbulbar region (g) Posterior part of male

the name Nemata has scientific preference, because no one can be in doubt as to included and excluded types. The corresponding vernacular name, nema, also has many other practical advantages including brevity and euphonious derivatives such as nematology and nematologist.

B. G. Chitwood included the Nemata or Nematoda (restricted sense) as a phylum in the series Aschelmintha, coordinate in rank with the Rotifera (Rotatoria), Gastrotricha, Kinorhyncha (Echinodera), and Nematomorpha but excluding the Acanthocephala, Platyhelmintha, and Nemerta. L. Hyman regarded the Platyhelminthes, Rhynchocoela (Nemertea), and Acanthocephala as phyla equivalent to the Aschelminthes; to these was added the class Priapulida. The Aschelminthes was not as compact as other invertebrate phyla, and for this specific reason Chitwood recognized the unsegmented worms as a subkingdom Scolecida or Amera and because of further questionable points in association it is now felt that grouping in series Parenchymata and Aschelmintha is not necessarily warranted. Some authorities presently consider as equivalent phyla the Platyhelmintha, Nemerta (or Rhynchocoela), Acanthocephala, Rotifera, Entoprocta, Gastrotricha, Kinorhyncha (or Echinodera), Nematomorpha, Nemata, and perhaps the Priapulida. Such matters as placement of groups at the ranks of series, phyla, classes, and orders are theoretic groupings and it is to be expected that there will be nearly as many differences of opinion as there are specialists working in the field.

Taxonomy. A classification of this group follows:

Phylum Nemata (Nematoidea, Nematoda)
- Class Secernentea (Phasmidia, Phasmidiea)
  - Subclass Rhabditia
    - Order Rhabditida
    - Order Strongylida
    - Order Ascaridida
    - Order Tylenchida
  - Subclass Spiruria
    - Order Spirurida
    - Order Camallanida
- Class Adenophorea (Aphasmidia, Aphasmidiea)
  - Subclass Chromadoria
    - Order Chromadorida
    - Order Monhysterida
  - Subclass Enoplia
    - Order Enoplida
    - Order Dorylaimida
    - Order Dioctophymatida

DIAGNOSIS OF THE PHYLUM

The Nemata are unsegmented or pseudosegmented worms with a basically circular cross section (cylindroid). The length of adults ranges from 150 microns to 8 meters; the external covering is a noncellular, simple to complex cuticle (not chitin or cuticulin), chiefly of scleroproteins. The anterior end bears sensory papillae or setae, usually six or three lips, and paired pseudolabia or jaws. The hypodermis (epidermis) is cellular to syncytial, usually enlarged medially and laterally as chords.

Musculature. The somatic musculature consists of four or more longitudinal bands (forming a muscular tube) of nonstriated, continuous, interlocking, spindle-shaped, unicellular muscle cells, each innervated by a peculiar protoplasmic innervation process to a medial nerve. Circular muscles are absent. The male has *specialized transverse* copulatory muscles in the caudal region and the female has diagonal muscles at the vulva. Dorsoventral and rectal muscles are commonly present. Additional special *somatointestinal muscles* are sometimes present. The body cavity is lined to a greater or lesser extent by pseudocoelomic membranes, and organs are suspended in the body cavity by mesenteries. The nuclei of all of these structures are apparently of limited and fixed numbers. The lining material originates as migratory mesenchymatous cells during embryogeny; protoplasmic processes from these cells tend to grow together, covering muscles and suspending organs. The body cavity is filled with colorless liquid and rarely shows tendencies toward subdivision; it is not a coelom, because of mesenchymatous rather than layer-split origin *See* CELL CONSTANCY.

Digestive tract. The digestive tract consists of an anterior, terminal oral opening, followed by a variously developed *stoma (basically a short cylindroid* tube), the esophagus (with its triradiate lumen,

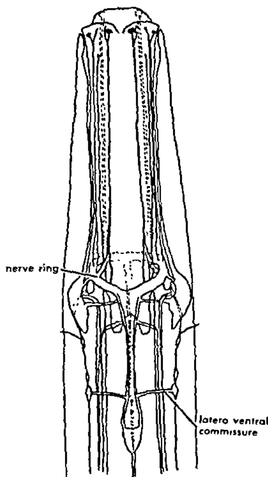


Fig 2. Nervous system of Spironoura affine

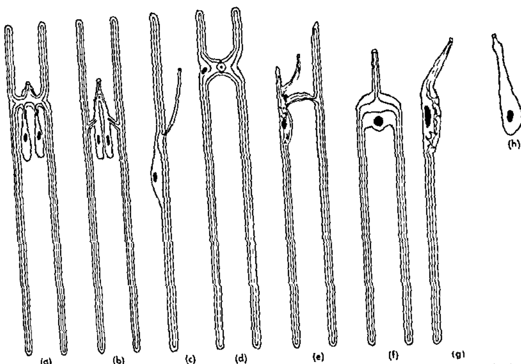


Fig 3 Representative types of excretory systems. (a) Rhabditoid type (b) Variant of rhabditoid type (c) Tylenchoid type (d) Oxyuroid type (e) Ascaroid type (f) Cephaloboid type (g) Anisokid (h) Single ventral cell type

one ray directed ventrad), esophagointestinal valve, intestine, rectum or cloaca, and ventral anus or cloacal opening The area from the stoma through the esophagointestinal valve is termed the foregut It is essentially a muscular organ of ingestion, with transverse radial muscles, and lined with cuticle formed by nuclei (termed marginal nuclei) at the ends of the esophageal radii The esophagus is variously modified into sections by means of bulbs and valves permitting suction and ingestion of food The esophagointestinal valve is fundamentally a set-off region of esophageal tissue connecting with the intestine Within the wall, or sometimes projecting from it, are three or more salivary (esophageal) glands, which may open into the stoma or at various places along the length of the esophagus A complex but very minute esophago-sympathetic nervous system serves to coordinate functions in the esophagus The midgut, or intestine is a straight tube of one cell layer, rarely with one or more ceca The interior edge of the cells may or may not form a striated border called the bacillary layer Posteriorly the intestine empties into the hindgut or rectum through an intestino-rectal valve Special unicellular rectal glands are commonly present The rectum and cloaca are lined with cuticle which has a distinct single layer of few cells The anus is posterior ventral, rarely subterminal to terminal.

Nervous system. The conspicuous part of the nervous system consists of a large anterior commissure, the nerve ring, which surrounds the esophagus and is connected with six cephalic papillary nerves, two lateral amphidial nerves and associated ganglia, and posterior somatic nerves and ganglia. The body has two primary motor nerves extending posteriorly in the two median chords In addition, the ventral nerve is actually a partly paired, partly single, ganglionated nerve chain without distinctively set-off ganglia, containing many associational cells There are hypodermal commissures to the lateral chords and lateral nerves innervating various sublateral tactile organs The ventral nerve is the chief nerve of the body and gives off nerve branches to several organs, including various parts of the digestive tract and copulatory organs (Fig 2)

The excretory system, of 1–4 cells, opens ventrally, usually in the esophageal region. The system may be branched, tubular, and extensive, a simple unicellular gland cell, rudimentary; or apparently absent A special tubular auxiliary excretory system of unknown significance is rarely present Flame cells are absent or have not been proven to exist

Reproductive system. Sexes are separate. Gonads are paired or single (very rarely multiple), tubular, with germ cells originating at the blind end or along the sides of the gonad. The sex opening in the female (and female-appearing hermaphrodites) is separate, on the ventral side of the body between the excretory pore and anus. Hermaphroditic females are peculiar in that both sperm cells and oocytes are formed at separate times from the same primordial germ cells See PROTANDRY; PROTOGYNY.

Male gonads open posteriorly into the ventral side of the rectum, forming a cloaca. One or two special sclerotized cuticular nonhollow spicules form in the dorsal wall of the cloaca. These spicules have special muscles and are extruded into the vagina of the female during copulation, presumably aiding in the transfer of sperm but not as a true penis or intromittant organ. Males have various modifications of cuticular sensory papillae and muscles, which aid in the copulatory act. Sperm may be of various types from flagellate to ameboid.

Females usually are oviparous; sometimes they are ovoviviparous with the larva hatching from the egg before deposition. The egg shell is chitinous, commonly covered externally with an albuminoid coat and internally by a waxy vitelline membrane. There are no nurse cells within the egg, but a zygote contains stored food in the form of protein. Fatty and glycogenous globules are rather uniformly distributed.

Cleavage of the embryo is bilateral and determinate, and the embryo takes a gross form more or less similar to the adult, hatches, and increases in cell number as it grows. These stages are marked by four castings of the cuticle, called molts (one or two molts may be within the egg shell), but there is no complete reconstitution of internal organs and therefore no true metamorphosis. However, there may be some rather striking changes in outline and actual structure of certain organs, particularly of the foregut. All stages lack the power of regeneration. *See* CLEAVAGE, EMBRYONIC.

Secernentea. In this class the primary excretory system consists of intracellular tubular canals, usually situated in the lateral chords, joined anteriorly and ventrally in an excretory sinus into which two ventral excretory gland cells may also open. There is usually a well-developed sinus nucleus, a nucleus in each subventral gland, if such are present, and a nucleus in the wall of the terminal excretory duct. In rare instances, in the females of some rhabditids there is in addition a paired auxiliary tubular excretory system (sometimes branched) also situated in the lateral chords which opens laterally posterior to the vulva. The amphids are usually small, externally porelike, and situated on the lips. Paired lateral cervical sensory papillae, deirids, are often present near the nerve ring. Three to six rectal glands usually open into the anterior part of the intestine. Caudal and hypodermal glands are absent, the male often has caudal cuticular alae at the tail, with paired rows of genital sensory papillae but without preanal glandular supplementary organs. Scattered or sublateral sensory organs usually are absent. Cuticular projections on the body in the form of cuticular setae are lacking and cephalic sensory organs usually are papilloid. Paired lateral caudal pores, phasmids (chemoreceptors), usually are present and well developed, though sometimes difficult to observe. These are apparently absent in parts of such large groups as the order Tylenchida.

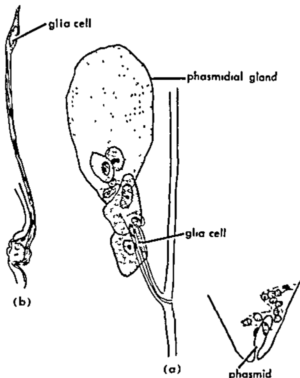


Fig. 4. (a) *Rhabditis terricola*, longitudinal section, female showing phasmidial gland. (b) *Spironoura affine*, longitudinal reconstruction of phasmids, showing phasmidial glands, neurons, and glia cells.

Adenophorea. In the class Adenophorea the excretory system, if present, consists of a single ventral cell in the body cavity. Tubular extensions in the chords are absent, and the terminal excretory duct is rarely lined with cuticle. Hypodermal glands are commonly well developed, and open through simple lateral or sublateral pores, or through tubular gland setae. Amphids are enlarged externally and the variously modified forms may be a combination unispire-pocket; a multispiral, circular, shepherd's crook, transversely lenticular; a simple pocket; or a small pore. Cephalic tactile sensory organs are setose to papilloid; somatic setae or papillae are common. Rectal glands are not proved. Adhesive caudal glands (usually 3) commonly open through a terminal valve, the spinneret, and act as a means of attachment for most aquatic forms to their substrates, but are absent in all obligate parasites. Males rarely have caudal alae but commonly have one or two series of preanal ventral supplementary organs which may be tactile or heavily sclerotized and may act as the orifice of adhesive glands to aid in copulation. Paired lateral caudal pores (phasmids) are absent, although corresponding unmarked areas termed phasma may be visible on the tails of some forms.

LIFE CYCLE

Life cycles among nematodes are diversified, however, basically, an adult female is inseminated at copulation by a male. The sperm finds the mature oocyte and enters. Thereafter an egg shell

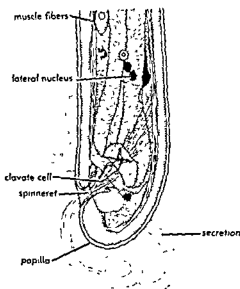


Fig 5 *Mononchulus ventralis* Tail showing details of spinneret.

(chitin) is formed by cell inclusions of the ovum, and a vitelline membrane (waxy) is formed inside the shell and ...

embryo undergoes cleavage, passing through a tad pole stage before assuming the eel-like larval form A total of four molts ...

utilized An alternation of generations of grossly dissimilar organisms rarely occurs.

BIONOMICS

Nematodes are one of the most successful of all groups of animals, because in any location where any other type of life exists (except microplankton), some stages of nemas, usually a multitude of individuals and kinds in full reproductive activity, are found Most of the important parasites of domestic animals and man are included in the class Secernentea, orders Rhabditida Strongylida, Ascaridida, Camallanida, and Spirurida The nemic parasites of plants, most of which are also included in the class Secernentea order Tylenchida may also be of great economic importance. A minority of the parasites of domestic animals and man, as well as those of plants, are included in the class Adenophorea The important adenophore parasites of domestic animals and man are included in the orders Dorylaimida and Dioctophymatida, whereas the minority of the important parasites of plants are also included in the order Dorylaimida. The majority of aquatic nemas are included in the same class in the orders Chromadorida, Monhysterida, and Enoplida

Many nemas are termed free living, but do eat products of decay, fungi, algae, and are economically unimportant animals All degrees of association exist with the various plant and animal groups.

A few specialized forms such as *Litomosoides carinii* are known to require free oxygen as adults (obligate aerobes), but the vast majority of the parasitic nemas studied have been found to be capable of prolonged periods of life without free oxygen (facultative anaerobes). None have been proved to be capable of completing their entire life cycle, including embryonic development, anaerobically. Glycogen is one of the chief energy sources, although stored fatty materials apparently provide the chief energy source in a few groups such as in the order Tylenchida. [B G C]

Bibliography: B G. Chitwood et al, *An Introduction to Nematology*, rev. ed., 1950——; B G. Chitwood, The English word "nema" revised, *Syst. Zool.*, 6(4):184–185, 1958; B. G. Chitwood, The designation of official names for higher taxa of invertebrates, *Bull Zool Nomenclature*, 15(25/28):860–895, 1958; E. C. Dougherty, Notes on the naming of higher taxa with special reference to the Phylum (or Class) Nematoda, *Bull. Zool. Nomenclature*, 15(25/28):896–906, 1958; L. H. Hyman, *The Invertebrates*, vol. 3, 1951.

Nematomorpha

A class of worms in the phylum Aschelminthes, ... lied ... in ... sitic in arthropods. The nematomorphs are found all over the world. They are divided in two orders, the Gordioidea and Nectonematoidea, with a total of 225 species

Morphology. The body is long and slender with a maximum length of 1.5 m and a diameter of 0.5–3 mm The females are longer than the males. The anterior end is rounded with a dark pigmented ring and a terminal mouth. The posterior end may be rounded with a terminal ...

body wall consists of three layers: an outer, rather thick fibrous cuticle; an epidermis consisting of a single layer of cells; and innermost, a muscle layer with longitudinal fibers only The surface of the cuticle may be smooth, or rough with rounded or polygonal thickenings called areoles. These may be flat or may form projecting structures, sometimes with bristles, and they may be perforated by pores and canals Between the areoles run interareolar furrows, often with wartlike structures and bristles. Special natatory bristles are developed in *Nectonema*. ...

Body cavi ... the body It ... minor spaces ... around the digestive system and the gonads. The digestive tube is always more or less degenerate, and the anterior part is often a solid string of cells. Ingestion of food is impossible, and the intact part of the digestive system

Male gonads open posteriorly into the ventral side of the rectum, forming a cloaca. One or two special sclerotized cuticular nonhollow spicules form in the dorsal wall of the cloaca. These spicules have special muscles and are extruded into the vagina of the female during copulation, presumably aiding in the transfer of sperm but not as a true penis or intromittant organ. Males have various modifications of cuticular sensory papillae and muscles, which aid in the copulatory act. Sperm may be of various types from flagellate to ameboid.

Females usually are oviparous; sometimes they are ovoviviparous with the larva hatching from the egg before deposition. The egg shell is chitinous, commonly covered externally with an albuminoid coat and internally by a waxy vitelline membrane. There are no nurse cells within the egg, but a zygote contains stored food in the form of protein. Fatty and glycogenous globules are rather uniformly distributed.

Cleavage of the embryo is bilateral and determinate, and the embryo takes a gross form more or less similar to the adult, hatches, and increases in cell number as it grows. These stages are marked by four castings of the cuticle, called molts (one or two molts may be within the egg shell), but there is no complete reconstitution of internal organs and therefore no true metamorphosis. However, there may be some rather striking changes in outline and actual structure of certain organs, particularly of the foregut. All stages lack the power of regeneration. See CLEAVAGE, EMBRYONIC.

Secernentea. In this class the primary excretory system consists of intracellular tubular canals, usually situated in the lateral chords, joined anteriorly and ventrally in an excretory sinus into which two ventral excretory gland cells may also open. There is usually a well-developed sinus nucleus, a nucleus in each subventral gland, if such are present, and a nucleus in the wall of the terminal excretory duct. In rare instances, in the females of some rhabditids there is in addition a paired auxiliary tubular excretory system (sometimes branched) also situated in the lateral chords which opens laterally posterior to the vulva. The amphids are usually small, externally porelike, and situated on the lips. Paired lateral cervical sensory papillae, deirids, are often present near the nerve ring. Three to six rectal glands usually open into the anterior part of the intestine. Caudal and hypodermal glands are absent; the male often has caudal cuticular alae at the tail, with paired rows of genital sensory papillae but without preanal glandular supplementary organs. Scattered or sublateral sensory organs usually are absent. Cuticular projections on the body in the form of cuticular setae are lacking and cephalic sensory organs usually are papilloid. Paired lateral caudal pores, phasmids (chemoreceptors), usually are present and well developed, though sometimes difficult to observe. These are apparently absent in parts of such large groups as the order Tylenchida.

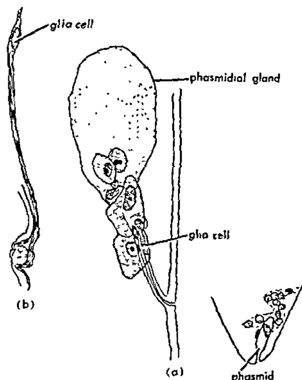


Fig. 4 (a) Rhabditis terricola, longitudinal section, female showing phasmidial gland. (b) Spironoura affine, longitudinal reconstruction of phasmids, showing phasmidial glands, neurons, and glia cells

Adenophorea. In the class Adenophorea the excretory system, if present, consists of a single ventral cell in the body cavity. Tubular extensions in the chords are absent, and the terminal excretory duct is rarely lined with cuticle. Hypodermal glands are commonly well developed, and open through simple lateral or sublateral pores, or through tubular gland setae. Amphids are enlarged externally and the variously modified forms may be a combination unispire-pocket, a multispiral, circular, shepherd's crook, transversely lenticular; a simple pocket, or a small pore. Cephalic tactile sensory organs are setose to papilloid; somatic setae or papillae are common. Rectal glands are not proved. Adhesive caudal glands (usually 3) commonly open through a terminal valve, the spinneret, and act as a means of attachment for most aquatic forms to their substrates, but are absent in all obligate parasites. Males rarely have caudal alae but commonly have one or two series of preanal ventral supplementary organs which may be tactile or heavily sclerotized and may act as the orifice of adhesive glands to aid in copulation. Paired lateral caudal pores (phasmids) are absent, although corresponding unmarked areas termed phasma may be visible on the tails of some forms.

LIFE CYCLE

Life cycles among nematodes are diversified, however, basically, an adult female is inseminated at copulation by a male. The sperm finds the mature oocyte and enters. Thereafter an egg shell

The family Conidae contains the poisonous cone shells which have caused several deaths by their poisonous bite *See* GASTROPODA. [W.J.C.]

Neognathae

One of the three superorders comprising the subclass Neornithes of the class Aves. All living birds are included in this subclass, as are all known fossil birds back through the Eocene. The Odontognathae differ from the Neognathae in possessing teeth, and the Ichthyornithes in possessing biconcave vertebrae. The Ratites, or large running birds, and the penguins are placed in the Neognathae, although both of these groups have previously been considered separate superorders. *See* AVES [C.B.C.]

Neomycin

Neomycin is a colorless antibiotic produced by certain strains of *Streptomyces fradiae* in submerged culture (*see* STREPTOMYCETACEAE). The fermentation medium usually consists of protein rich materials such as soybean, cottonseed meal, or peanut meal, together with some sugar or sugar rich materials, such as distiller's solubles, as well as certain supplementary minerals. The length of the fermentation [illegible]

[illegible] stance in the filtered broth is then absorbed, either on activated carbon or on ion-exchange resins (*see* ION EXCHANGE) The neomycin is eluted, or removed, from the carbon with 10% acetone at pH 2 Commercial production of neomycin in the United States in 1956 was 17,000 lb.

The antibiotic is a polybasic, water-soluble substance first isolated by S A Waksman and H. A Lechevalier in 1948 The neomycin complex was found to consist of two isomeric substances, now recognized as neomycin B and neomycin C. Commercial neomycin is composed mainly of neomycin B and is usually in the form of a sulfate, a white amorphous powder An A fraction was at first also described, it was later found to be a degradation product comparable to neamine

Chemically, the structure of neomycin ($C_{23}H_{46}N_6O_{13}$) is analogous to that of streptomycin It consists of two fractions neobiosamine (diamino-hexose + pentose) and neamine (diamino hexose + 2 desoxystreptomine) Several closely related compounds notably catenulin and framycetin have also been isolated as metabolic products of different species of *Streptomyces*.

Neomycin is active against a great variety of gram positive cocci and rods gram-negative rods, and acid fast bacteria, notably the tuberculosis organism (*see* TUBERCULOSIS) It is not active against anaerobic bacteria, fungi, most protozoa, rickettsiae, and viruses (*see* RICKETTSIOSES, VIRUS) Development of resistance to neomycin is slower than that to streptomycin

The LD_0 for subcutaneous injection of neomycin sulfate in mice is 125 mg/kg, and the LD_{50}, or dosage at which 50% of the animals tested died, is 165–250 mg/kg (*see* LETHAL DOSE 50). The LD_{50} for intraperitoneal administration is about the same as that for subcutaneous injection. Intravenously, the neomycin preparations are about five times as toxic. Orally, the LD_{50} is greater than 2800 mg/kg.

When given by injection, neomycin is well distributed in the body fluids. When given orally, little is absorbed through the intestinal wall. When given in high enough concentration for a [illegible] mycin is markedly nontoxic, nonirritating, and has a low index of allergenicity.

Clinically, neomycin is used topically in the treatment of bacterial infections of the skin and orally in the treatment of bacterial infections of the digestive tract. It can also be used systemically in the treatment of deep-seated bacterial infections that are resistant to other, less toxic antibiotics.

The chemotherapeutic uses of neomycin are largely twofold: (1) in the treatment of bacterial infections of the skin, eyes, and oronasal cavity; (2) in enteric bacterial infections, such as infantile diarrheal diseases, and also as a general intestinal antiseptic.

It is also used in the treatment of peritonitis, in surgery and wound infections, in nonspecific urethritis, as an aerosol, and in veterinary medicine. Neomycin is often used in combination with bacitracin or gramicidin (*see* BACITRACIN; TYROTHRICIN)

In ophthalmology, neomycin is also used in combination with a steroid, cortisone. For topical application, advantages of neomycin are: great efficacy in [illegible] organisr[illegible] irritanc[illegible] the skin [illegible] wound, and great stability. Should resistance develop in organisms that might [illegible] toxic systemically than neomycin, can be used subsequently if necessary.

As an intestinal antibiotic the advantages of neomycin are its wide antimicrobial spectrum, low toxicity due to limited absorption, chemical stability in the presence of the digestive enzymes and food, rapidity of action, limited development of bacterial resistance, and noninterference with tissue growth and repair Neomycin is given orally [illegible] combat [illegible] [illegible] case of many [illegible] antibiotics, in a stimulation of the fungal flora, either of the skin or of the bowels

For veterinary medicine special neomycin ointments are available as for the treatment of bovine mastitis *See* INDUSTRIAL MICROBIOLOGY. [S.A.W.]

seems to be adapted for excretory functions. During the parasitic stage, food is obtained through the body surface by means of digestive enzymes.

Nervous system. This consists of a cerebral mass lying ventrally in the head and a ventral nerve cord which originates in the epidermal layer. Little is known of the sensory organs. Probably the bristles and warts of the cuticle have sensory functions. A rudimentary eye is found in the genus *Paragordius*.

Reproduction. The sexes are always separate and the gonads are paired and stringlike extending the length of the body. In males, the gonads are connected with the cloaca by sperm ducts. In females, the ovaries form a large number of lateral diverticula in which the eggs ripen The oviducts enter the cloaca separately. A sac, the seminal receptacle, extends anteriorly from the cloaca.

During copulation the male coils itself around the female and places a drop of sperm near the cloacal opening of the female The sperm cells actively enter the seminal receptacle. The eggs are laid in water in strings and the adults die after egg laying. When hatched, the larvae swim to an aquatic arthropod. They penetrate the body wall of the host by means of their characteristic proboscis, which is armed with hooks and three long stylets. The larvae of some species may secrete a special mucus in which they encyst until they are accidentally ingested by the right host, which may be a terrestrial insect. The gradual development in the host lasts some months without any metamorphosis. When mature, the worms leave the host. *See* GORDIOIDEA, NECTONEMATOIDEA. [B.J.MU.]

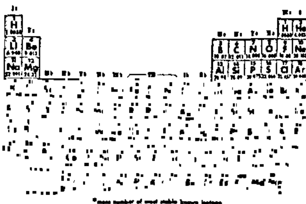
Nematophytales

Plants from the Silurian and Devonian periods that bear some resemblance to the brown seaweeds (Phaeophyta), but cannot be placed in any living group *Prototaxites* was a large plant with a trunklike body that sometimes attained a diameter of 3 ft. The trunks were constructed of intertwined tubes of two sizes, small branched ones 5–10 microns in diameter and large unbranched ones up to 50 microns Reproductive organs are unknown *Nematothallus* was a thin leaflike growth of tubular construction with a cutinized outer surface. Within some of the tubes small spores borne in tetrads have been observed. *See* PALEOBOTANY. [C.A AR]

Neodymium

Element number 60, neodymium, Nd, is a metallic element belonging to the rare-earth group Its atomic weight is 144.27, and the naturally occurring element includes the stable isotopes Nd^{142} 27.11%, Nd^{143} 12.17%, Nd^{144} 23.85%, Nd^{145} 8.30%, Nd^{146} 17.22%, Nd^{148} 5.73%. It was discovered by C. F. Auer von Welsbach in 1885 when he separated the so-called element didymium into two fractions, neodymium and praseodymium The oxide, Nd_2O_3, is a light blue powder. It dissolves in mineral acids to give reddish-violet solutions. For properties of the metal, *see* RARE-EARTH ELEMENTS

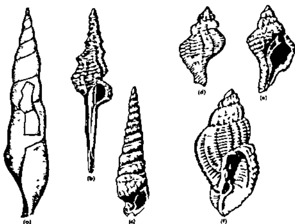
The metal slowly oxidizes in air at room temperatures and is slowly attacked by cold water The salts have found application in the ceramic industry for coloring glass and for glazes. The glass



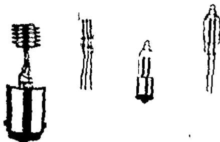
made with didymium is particularly useful in goggles used by glass blowers, since it absorbs the intense yellow D line of sodium present in the flame. [F.H SP.]

Neogastropoda

An order of gastropods, also known as the Stenoglossa, which contains the most highly developed snails. Respiration is by means of ctenidia. The nervous system is concentrated, an operculum is usually present, and the sexes are separate. All families in this order are marine. The family Muricidae contains the rock snails, all predatory, and many species in this family are important economically because they feed on oysters. The family *Buccinidae*, most abundant in northern seas, contains the whelks in the genus *Buccinum* which are frequently used for fish bait. Other families such as the Volutidae, Olividae, and Harpidae are prized by shell collectors for their beautiful coloration



Neogastropoda. (a) An Ordovician species of *Subulites*, a widespread Ordovician and Silurian genus (b) *Falsifusus*, an early Tertiary genus (c) A Miocene species of the familiar existing *Terebra* (Tert.-Recent) (d, e) *Urosalpinx* (Tert.-Recent) (f) *Cancellaria* (Tert.-Recent). (R. R. Shrock and W. H. Twenhofel, *Principles of Invertebrate Paleontology*, 2d ed., McGraw-Hill, 1953)



Typical glow lamps

conventional bases that vary with the size of the lamps (see illustration)

Operation. Electrodes sealed in a neon atmosphere will emit electrons if a sufficient voltage difference is impressed across them In glow lamps the electrodes are usually treated to emit electrons freely With a sufficiently high voltage between electrodes, the velocity of electron flow is high enough to ionize the neon nearest the *negative* electrode (cathode) The neon then emits a reddish-orange glow similar to the color of neon sign tubing With direct current the glow is restricted to the immediate vicinity of the negative electrode. With alternating current, both electrodes act alternately as cathodes, and the glow appears alternately at both surfaces. At usual frequencies, the alternations occur so rapidly that both electrodes appear to glow constantly

In dc circuits, the voltage across the electrodes may be reduced significantly once the lamp has started without causing the lamp to go out Direct-current starting voltages for typical glow lamps range from 65 to 90 volts, while the minimum operating voltage at which the glow will be maintained may be 10–15 volts lower On alternating-current circuits, the maintaining voltage is nearly the same as the starting voltage.

This glow discharge is like the arc in a vapor lamp in that its resistance decreases with increasing current Therefore, a current-limiting element must be used in the electric circuit to maintain a desired stabilized current in the circuit Because the current in glow lamps is usually a few milliamperes or less, it is both practical and economical to use a small resistor as a ballast The larger glow lamps with screw bases have resistors in the bases, smaller lamps require an external ballast resistor. The appropriate resistance value depends on the lamp type

Applications. The neon glow lamp is inherently a low wattage source that produces light at relatively low efficiency when compared to filament lamps and other vapor lamps Its lighting applications are limited to those where a low power source is needed to provide an *indication* of the location or status of equipment. These include illuminated wall switches in homes and indicator lights on the panels of electrical devices.

In electronic circuits involving relatively low power, neon lamps are used in many ways. They are used in counter and memory elements of computers, in voltage regulators, in relaxation oscillators, and in trigger circuits to operate relays and similar devices These applications are practical because of the lamp's unique electrical characteristics, its small size, and its light weight.

Characteristics. The useful life of a glow lamp is not terminated by a burnout, as is the life of lighting lamps, but by a gradual rise in starting and maintaining voltage and blackening of the inner walls of the bulb, reducing light output. If the lamp is used as an indicator light, the reduction in brightness will determine its life, which may be 5000–25,000 hours, depending on the application.

When the lamp is used as a circuit element and

... burning hours Maintaining voltage rises about half as fast as starting voltage.

External factors may also affect the operating characteristics of glow lamps. The sensitized electrodes of glow lamps release electrons in the presence of light. In total darkness, the starting voltage may be 100 volts higher than in light. In totally dark enclosures, slight amounts of light or other *radiation*, or *electrostatic* fields, may be employed to overcome the dark effect. Darkness does not affect maintaining voltage. *See* VAPOR LAMP.

[A MA.]

Bibliography· General Electric Company, *Glow Lamps as Circuit Control Components 3-6177-R*, General Electric Company, *Lamp Bulletin LD-1*; Illuminating Engineering Society, *IES Lighting Handbook*, 2d ed, 1952, Stanford University, *Neon Lamps as Circuit Elements*, Rept. 10.

Neoplasia

Neoplasia is a form of growth producing cells similar to those normally found in the body but usually less differentiated in type and less organized in structure. The terms neoplasm and tumor are used synonymously.

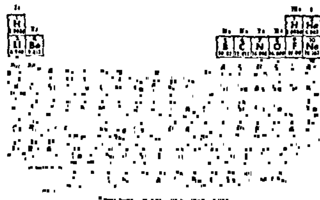
Although a differentia...
hyperplas...
basis, an ...
The cell ...
eventuate ...
and the f...
functional ...
of neopla... intermitotic and do not mature or differentiate In the former case, potentialities have reached full expression, while in the latter the possibility of further development is always present.

On transplantation, the hyperplastic focus behaves exactly as does a normal, adult tissue. Conversely, the transplantation reactions of an area of

Neon

Chemical element number 10, neon, Ne, is a member of the family of noble gases, group 0 of the periodic table (*see* INERT GASES). The only commercial source of neon is the earth's atmosphere, although traces of neon are found in minerals and meteorites.

Uses. When neon at low pressure is excited in an electric discharge, it emits a brilliant orange-red light; this property accounts for the largest use of neon, the filling of the electric neon signs. To obtain colors other than the characteristic orange-red of neon, other inert gases and mercury vapor may be added.



Neon is used as the current-carrying agent in lightning arrestors, virtually no current is carried at voltages below the breakdown potential of the neon, but when lightning strikes, the neon is ionized and allows the current to flow to ground Neon discharge tubes are used as overload protection for some electric motors to guard them against damage from surges in current.

Neon is also used in some kinds of electron tubes, in Geiger-Muller counters, in spark-plug test lamps, and in warning indicators on high-voltage electrical lines. A very small wattage produces visible light in neon-filled glow lamps; such lamps are used as economical night and safety lights

Because the boiling point of neon is only 27K° above that of liquid hydrogen, liquid neon may be used as a nonflammable substitute for liquid hydrogen in producing low temperatures.

Occurrence. Neon constitutes 18 18 parts per million by volume of the earth's atmosphere, and this neon is a mixture of 3 stable isotopes: 90 92% volume neon-20, 0.26% neon-21, and 8.82% neon-22. No naturally occurring radioactive isotopes of neon are known.

It is estimated that about 5×10^{-7}% by weight of the earth is neon. Neon also occurs outside the earth; the best estimate is that there are 8.6 times as many atoms of neon as of silicon in the visible universe.

Discovery. Neon was discovered in England in 1898 by Sir William Ramsay and M. W. Travers who found it in the most volatile portion of the mixture of inert elements left after oxygen and nitrogen had been chemically removed from air. The fact that a new element was present was ascertained by the discovery of new lines in the emission spectrum of the residual gas.

Physical properties of neon

Property	Value
Atomic number	10
[illegible]	20 183
[illegible]	−248 8
[illegible]	−246 0
[illegible] g/liter	0 8999
Liquid density at its boiling point, g/ml	1 2
Solubility in water, ml gas/1000 g water at 25°C and 1 atm pressure	11 1

Properties. Neon is colorless, odorless, and tasteless; it is a gas under ordinary conditions Some of its other properties are given in the table

Neon does not form any chemical compounds in the ordinary sense of the word, and there is only one atom in each molecule of neon

Production. In the production of neon from air, the air is first liquefied. A small amount of it remains uncondensed; this contains the hydrogen, helium, and neon, with a little nitrogen and oxygen. Most of the nitrogen and oxygen is removed by low-temperature adsorption on activated carbon. The hydrogen is burned to water and the residual gas is dried The remaining gaseous impurities are then separated by selective adsorption on activated carbon at carefully regulated temperatures and pressures.

Analytical methods. The principal modern methods of detecting and quantitatively determining the neon content in gases are mass spectrometry and gas chromatography. Until these methods were developed, it was necessary to separate neon from other inert gases by selective low-temperature adsorption on activated carbon in order to determine how much neon was present in a mixture The older method of detecting neon is by its characteristic emission spectrum, obtained by passing a gas sample through an electric discharge tube at low pressure and analyzing the light with a spectrometer. *See* ATMOSPHERIC GASES, PRODUCTION OF, VAPOR LAMP. [C S.C.]

Bibliography: F. P. Gross, Jr., Rare gases in everyday use, *J. Chem. Educ.*, 18:535–539, 1941; R F. Kirk and D. F Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 7, 1951; F. J. Metzger, Traces [of rare gases] from tons, *Ind Eng. Chem*, 27:112–116, 1935, H A. Miller, *Luminous Tube Lighting*, 2d ed., 1947; S. C Miller, *Neon Signs and Cold-Cathode Lighting*, 2d ed., 1952.

Neon glow lamp

A low-wattage lamp often used as an indicator light or as an electronic circuit component The neon lamp usually consists of a pair of electrodes sealed within a bulb containing neon gas at a low pressure. Some of the smaller bulbs are equipped with wire leads that are connected directly into the electrical supply circuit; others are equipped with

feldspar, the rock may be called nepheline monzonite. Nepheline syenites have many features in common with syenites into which they grade, but chemically, mineralogically, and texturally they are much more variable.

Composition. The alkali feldspar is soda-rich and usually exhibits perthitic texture (intergrown potash and soda feldspars). Barium-rich feldspar cores may be surrounded by barium-poor shells to give a zonal structure. If plagioclase occurs as discrete grains, it is usually albite or sodic oligoclase. More calcic types are found in nepheline monzonite. In some rock types the perthite is rimmed by albite.

Nepheline, a gray mineral with a greasy appearance, is also a major constituent. It may be highly altered and is commonly converted to bright yellow cancrinite. Additional feldspathoids (including sodalite, analcite, and leucite) may occur in minor quantities. See FELDSPATHOID.

Biotite mica rich in iron and titanium frequently shows a zonal structure with light-colored cores fringed by darker borders. The amphiboles are usually soda-rich (arfvedsonite, hastingsite, and riebeckite), but zoned brown hornblende occurs in some varieties. Pyroxenes are also soda-rich and show zoning with more diopsidic cores and aegirine-augite or aegirite borders. Some crystals show hourglass structure.

The most common accessory minerals are sphene, zircon, apatite, ilmenite, and magnetite. Rarely fluorite, garnet, corundum, and a variety of unusual minerals may be present.

Texture and structure. Equigranular texture (uniform grain size) is most common, and locally very coarse (pegmatitic) phases occur. Porphyritic texture (large crystals in finer-grained matrix) is almost confined to the nepheline syenite porphyries. The phenocrysts, where present, however, are usually [illegible] (with [illegible] textu[illegible] tal ou[illegible] line [illegible] large alkali feldspar grains enclose nearly euhedral crystals of nepheline. Various combinations of graphic (cuneiform) intergrowths occur between alkali feldspar, feldspathoids, and mafics. Tabular feldspar crystals and streaks of mafic minerals may show subparallel arrangement producing a flow structure in the rock.

Occurrence. Nepheline syenite and related rocks are rare and generally occur in small bodies (dikes, sills, laccoliths, stocks, and small irregular plutons). Only a few large bodies are known. Three of the largest are in southern Greenland, Pilaansberg in South Africa, and Kola Peninsula, U.S.S.R. Nepheline syenites may be associated with alkali syenites, with other feldspathoidal rocks, or with some alkali granites.

Formation. The origin of nepheline rocks is still a much debated problem. Many of these rocks may have formed from magma (molten rock material) of nepheline syenitic composition. This magma may have been derived from a basaltic one by fractional crystallization in which certain early-formed crystals were removed from the melt. By assimilation of abundant limestone, certain magmas may be desilicated to yield nepheline syenites. Some nepheline syenites and related rocks are believed to be of metamorphic and metasomatic origin. They may have formed from other rock types by introduction of certain elements and removal of others. See IGNEOUS ROCKS; MAGMA; METASOMATISM. [C.A.CA.]

Nephelinite

A dark-colored, aphanitic (very finely crystalline) rock of volcanic origin, composed essentially of nepheline and pyroxene.

The texture is usually porphyritic with large crystals (phenocrysts) of augite and nepheline in a very fine-grained matrix. Augite phenocrysts may be diopsidic or titanium-rich and may be rimmed with soda-rich pyroxene (aegirine-augite). Microscopically the matrix is seen to be composed of tiny crystals or grains of nepheline, augite, aegirite, and sodalite with occasional soda-rich amphibole, biotite, and brown glass. If leucite becomes the dominant feldspathoid, the rock becomes a leucitite. If calcic plagioclase exceeds 10%, the rock passes into tephrite and basanite. If olivine is present, the rock is an olivine nephelinite (nepheline basalt). Accessories usually include magnetite, ilmenite, apatite, sphene, and perovskite.

Nephelinite and related rocks are very rare. They occur as lava flows and small, shallow intrusives. A great variety of these feldspathoidal rocks is displayed in Kenya Colony, East Africa. See FELDSPATHOID, IGNEOUS ROCKS; LEUCITE ROCK.

[C.A.CA.]

Nephelometric analysis

A method of chemical analysis based on measurement of the cloudiness of solutions. In nephelometry, the intensity of the light scattered at right angles to the light beam by [illegible] ured, rather th[illegible] directly transm[illegible]

Simple nephelometers consist of a light source focused on the sample cell and a detecting element positioned at right angles to the light beam. Since a clear solution will scatter no light at right angles to the beam, nephelometry measures the difference

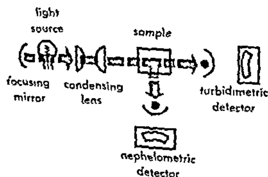


Diagram of nephelometric apparatus.

neoplasia are unique and unlike those of any other tissue state. Unlike normal adult tissue, a focus of neoplasia will not survive transfer to a normal unrelated individual and will grow only when transplanted back elsewhere in the body of the original host or to an individual bearing a neoplastic lesion of the same type. In other words, it is dependent for its continued existence on factors peculiar to the tumor-bearing individual and such factors are not supplied by normal individuals.

At this stage of development, the neoplasm is biologically dependent and clinically benign. However, with continued residence in the primary host, it attains independence of the factors concerned in genesis and development and gains the ability to grow in their absence. The attainment of autonomy or independence is associated with the attainment of the *ability to invade* and *metastasize*, and the neoplasm becomes clinically malignant.

A point to be emphasized is that the properties that make a tumor malignant are not present from the initiation of neoplasia but are developmental acquisitions. Cancer is not a sudden transformation of normal cells but, rather, the final stage of a developmental process. *See* ONCOLOGY. [W.F.P.]

Neornithes

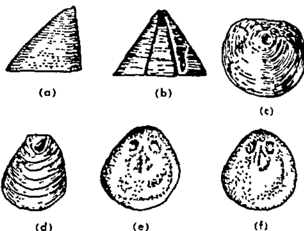
The subclass of Aves containing all known birds except *Archaeopteryx*, which alone constitutes the subclass Archaeornithes. The Neornithes are divided into three superorders: Odontognathae, Ichthyornithes, and Neognathae. A superorder Palaeognathae was formerly recognized for the Ratites, or large running birds, but this has been shown to be an artificial group based on convergent characters. The penguins are also placed in the Neognathae, although some authors segregate them as a superorder Impennes. *See* AVES; RATITES. [K.C.P.]

Neoteny

A phenomenon among some salamanders where larvae of large size, while still retaining the gills and other larval features, become sexually mature, mate, and produce fertile eggs. In certain lakes of Mexico, only the neotenous larvae are present and are called axolotls. Neoteny occurs in certain species of the family Ambystomidae, especially in *Ambystoma tigrinum* of some localities, and commonly in the large *Dicamptodon ensatus* of the Pacific Coast. It also occurs in some Texas and Oklahoma species of *Eurycea*. Neotenous larvae of *A. tigrinum* can be made to metamorphose to adult form if treated with thyroid extract. *See* PAEDOGENESIS. [T.I.S.]

Neotremata

An order of the class Inarticulata, in which the animals have round or elliptical bivalve shells. The pedicle emerges from a hole in the ventral valve and may be lost when the ventral valve adheres to the substratum. The pedicle opening may be absent due to atrophy of the pedicle. The shell substance may be chitinous, chitinophosphatic, or calcareous



(a) Lateral, (b) posterior, and (c) top views of a ventral valve of *Acrotreta gemma* (Ordovician, Newfoundland); (d) ventral, (e) exterior, and (f) interior of *Crania antiqua* (Cretaceous) (*From W. H. Twenhofel and R. R. Shrock, Principles of Invertebrate Paleontology, McGraw-Hill, 1953*)

Homoedeltidia and pseudochilidia are usually not well developed. Generally, the exterior exhibits coarse concentric growth lines. The shell is flattened to conical. The Neotremata arose from the Atremata. Of the four superfamilies, two, Discinacea and Craniacea, are extant. Living representatives have been recorded from shallow to abyssal depths (2737 fathoms) and the species are cosmopolitan. Their geological range extends from the Cambrian to Recent. *See* INARTICULATA. [K.H.]

Neper

A unit of attenuation used in transmission-line theory. On a uniform transmission line having waves traveling in only one direction, the magnitudes of voltage E and of current I decrease with distance x traveled, as given by the equation

$$\frac{E}{E_0} = \frac{I}{I_0} = \epsilon^{-\alpha x}$$

where E_0, I_0, and α are constants. The attenuation in nepers between the points where E_0 and I_0 are measured and where E and I are measured is

$$\alpha x = \ln \frac{E_0}{E} = \ln \frac{I_0}{I}$$

where ln denotes the natural (or Napierian) logarithm.

The word neper originated from a misspelling of the proper name Napier. One neper equals 8.686 decibels, the decibel being the practical unit of attenuation. *See* DECIBEL; TRANSMISSION LINES; TRANSMISSION THEORY AND METHODS. [E.W.K.]

Nepheline syenite

A phaneritic (visibly crystalline) plutonic rock with granular texture, composed largely of alkali feldspar (orthoclase, or microcline, usually perthitic), nepheline, and dark-colored (mafic) minerals (biotite, soda-amphibole, and soda-pyroxene). If sodic plagioclase exceeds the quantity of alkali

feldspar, the rock may be called nepheline monzonite. Nepheline syenites have many features in common with syenites into which they grade, but chemically, mineralogically, and texturally they are much more variable

Composition. The alkali feldspar is soda-rich and usually exhibits perthitic texture (intergrown potash and soda feldspars). Barium-rich feldspar cores may be surrounded by barium-poor shells to give a zonal structure If plagioclase occurs as discrete grains, it is usually albite or sodic oligoclase More calcic types are found in nepheline monzonite In some rock types the perthite is rimmed by albite

Nepheline, a gray mineral with a greasy appearance, is also a major constituent It may be highly altered and is commonly converted to bright yellow cancrinite. Additional feldspathoids (including sodalite, analcite, and leucite) may occur in minor [illegible]

[illegible] fringed by darker borders The amphiboles are usually soda-rich (arfvedsonite, hastingsite, and riebeckite), but zoned brown hornblende occurs in some varieties Pyroxenes are also soda-rich and show zoning with more diopsidic cores and aegirine-augite or aegirite borders Some crystals show hourglass structure.

The most common accessory minerals are sphene, zircon, apatite, ilmenite, and magnetite Rarely fluorite, garnet, corundum, and a variety of unusual minerals may be present

Texture and structure. Equigranular texture (uniform grain size) is most common, and locally very coarse (pegmatitic) phases occur *Porphyritic texture* (large crystals in finer-grained matrix) is almost confined to the nepheline syenite porphyries The phenocrysts, where present, however, are usually sanidine and mafics Feldspars may be anhedral (without crystal outline) to give allotriomorphic texture, or they may be nearly euhedral (with crystal outline) and associated with interstitial nepheline or mafics Poikilitic texture is found where large alkali feldspar grains enclose nearly euhedral crystals of nepheline Various combinations of graphic (cuneiform) intergrowths occur between alkali [illegible] felds[illegible] show [illegible] [illegible] structure in the rock

Occurrence. Nepheline syenite and related rocks are rare and generally occur in small bodies (dikes, sills laccoliths, stocks, and small irregular plutons) Only a few large bodies are known. Three of the largest are in southern Greenland, Pilaansberg in South Africa, and Kola Peninsula U.S.S.R. Nepheline syenites may be associated with alkali syenites, with other feldspathoidal rocks, or with some alkali granites

Formation. The origin of nepheline rocks is still a much debated problem Many of these rocks may have formed from magma (molten rock material) of nepheline syenitic composition This magma may have been derived from a basaltic one by fractional crystallization in which certain early-formed crystals were removed from the melt. By assimilation of abundant limestone, certain magmas may be desilicated to yield nepheline syenites. Some nepheline syenites and related rocks are believed to be of metamorphic and metasomatic origin. They may have formed from other rock types by introduction of certain elements and removal of others See IGNEOUS ROCKS; MAGMA; METASOMATISM. [C.A.CA.]

Nephelinite

A dark-colored, aphanitic (very finely crystalline) rock of volcanic origin, composed essentially of nepheline and pyroxene.

The texture is usually *porphyritic with large* crystals (phenocrysts) of augite and nepheline in a very fine-grained matrix. Augite phenocrysts may be diopsidic or titanium-rich and may be rimmed with soda-rich pyroxene (aegirine-augite). Microscopically the matrix is seen to be [illegible]

[illegible] glass. If leucite becomes the dominant feldspathoid, the rock becomes a leucitite. If calcic plagioclase exceeds 10%, the rock passes into tephrite and basanite. If olivine is present, the rock is an olivine nephelinite (nepheline basalt). Accessories usually include magnetite, ilmenite, apatite [illegible]

[illegible] in Kenya Colony, East Africa. See FELDSPATHOID; IGNEOUS ROCKS; LEUCITE ROCK.

[C.A CA.]

Nephelometric analysis

A method of chemical analysis based on measurement of the cloudiness of solutions. In nephelometry, the intensity of the light scattered at right angles to the light beam by a suspension is measured, rather than the decrease in the intensity of directly transmitted light as in turbidimetry

Simple nephelometers consist of a light source focused on the sample cell and a detecting element positioned at right angles to the light beam Since a clear solution will scatter no light at right angles to the beam, nephelometry measures the difference

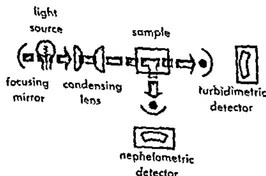


Diagram of nephelometric apparatus.

between no light and various amounts of scattered light, making it sensitive to the detection of small amounts of turbidity. Nephelometers with both photoelectric and visual detectors are used. The former are more sensitive and accurate than the latter.

Variables, such as particle size, stability of the suspension, and homogeneity of the turbidity, all of which affect turbidimetric analysis, also affect nephelometric methods and limit their precision and accuracy. Most analyses are carried out after the preparation of a standard curve using known mixtures whose composition approximates that of the sample.

The same types of systems may be analyzed by both nephelometry and turbidimetry. The advantage *of nephelometry is its greater sensitivity, accuracy,* and precision in the determination of small amounts of turbidity. Turbidimetry is somewhat simpler and better for more highly turbid solutions though the calibration curves generally deviate further from linearity than those of nephelometry. *See* OPTICAL METHODS OF CHEMICAL ANALYSIS; TURBIDIMETRIC ANALYSIS. [R.F.C.]

Nephritis

An inflammation of the kidney. Nephritis is a nonspecific term that refers more to the clinical findings in certain renal diseases but that may be correlated to inflammations primarily affecting certain areas of the kidney. No present system of classification is entirely adequate. Much confusion has resulted from the mixing of clinical entities and pathologic processes encountered in different disease states. Many renal diseases have an unknown or incompletely understood etiology. The *general clinical signs of kidney damage are simi*lar for many diseases because the kidney may react to damage in only a few ways. They include variable degrees of albuminuria, hematuria, retention of renal wastes, hypertension, and uremia.

Some clarification of the renal diseases has been attained by considering the portion of the kidney first or principally involved. On this basis, renal diseases can be categorized as glomerular, tubular, interstitial, and vascular, each in relationship to *various* kinds of inflammation (nephritis) or degeneration (nephrosis). There are few pure lesions.

Glomerulonephritis is primarily an inflammation of the capillary tufts of the nephron, or renal unit. Secondary or progressive changes may be induced in other areas. Acute, subacute, and chronic forms are recognized, as well as less-common circumscribed limited forms and hemorrhagic types.

The etiologic mechanism of acute glomerulonephritis appears to be some kind of sensitivity reaction to a bacterial product derived from prior infections with streptococcal organisms. Other causes of the subacute, chronic, and special forms include degenerations with accompanying inflammation, bacterial invasion, and emboli from various sources. *See* STREPTOCOCCUS.

Tubular nephritis may be present, but usually nephrosis is primary. Common causes include chemical poisons, certain toxic or metabolic states, and some forms of infection.

Interstitial nephritis occurs most often during severe generalized infections and either may produce *a diffuse general inflammation in the kidney or is* restricted to focal lesions. Pyelonephritis is a very common, often chronic, inflammation of both interstitial areas and the renal pelvis.

Vascular inflammations occur principally in the smaller arteries of the kidney. These may be due to a systemic vascular disease, such as arteriosclerosis, or to involvement of the vessels by inflammations which begin elsewhere. One of the most serious forms is that of malignant hypertension, a disorder of young or middle-aged adults. Rapidly appearing inflammatory lesions cause tissue death (necrosis) of the vessels with the production of hemorrhagic sites. *See* URINARY SYSTEM. [E.G.ST.]

Neptune

The outermost of the four giant planets. Its discovery, in 1846, within a degree from the position theoretically predicted was one of the great achievements of celestial mechanics. Difficulties in accounting for the observed motion of Uranus by the effect of the perturbations caused by the other planets known at the end of the eighteenth century led, early in the nineteenth century, to the suspicion that another unknown planet, beyond the orbit of Uranus, might be responsible for the unexplained perturbations. The difficult problem of deriving the mass and orbital elements of the unknown planet was solved independently in 1845–1846 by U. J. Leverrier in Paris and by J. C. Adams in Cambridge, England. *The optical search undertaken by* the English astronomers at Cambridge was still in progress when the planet was seen by J. G. Galle at Berlin on September 23, 1846, only 52′ away from the position predicted by Leverrier.

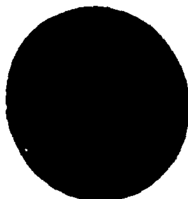
The planet and its orbit. The actual orbit of Neptune and its mass differ notably from the values predicted by Adams and by Leverrier, who had assumed for the mean distance of the planet to the Sun the value given by Bode's law, namely, 38.8 astronomical units, whereas it is only 30.1 units, or 2.8×10^9 miles. The eccentricity is only 0.009, the second smallest (after Venus) among the main planets; the sidereal revolution period is 164.8 years, the mean orbital velocity is 3.4 mi/sec; the inclination of the orbital plane to the ecliptic is 1°.8. *See* PLANET.

Neptune is not visible to the naked eye but can be seen through a small telescope as a greenish star of about eighth magnitude. Micrometer measurements with large telescopes give a mean apparent diameter of about 2″.0 at mean opposition (that is, when closest to Earth), corresponding to a mean linear diameter of 28,000 miles with an uncertainty of a few per cent. The polar flattening is not directly observable, but the ellipticity corresponding to the rotation is 0.02. The volume is

about 42 times that of the Earth The mass, 17 25

major planets The acceleration of gravity at the surface is about 15 m/sec² and the centrifugal force at the equator about 0 3 m/sec²

The apparent visual magnitude of the planet Neptune at its mean opposition is 7 6 and the corresponding value of the reflectivity (geometrical albedo) about 0 5, a value characteristic of the other major planets and indicating the presence of a dense, cloudy atmosphere A large telescope shows Neptune as a greenish disk, strongly darkened near the limb, on which vague irregular markings unlike those of Saturn and Uranus have occasionally been seen (see illustration) The rota-



Telescopic appearance of Neptune

tion period determined spectroscopically is about 15h 45m, with an uncertainty of perhaps 1h, and the direction of the motion is direct. The inclination of Neptune's equatorial plane to its orbit is about 29°

The optical spectrum of Neptune is marked by very strong absorption bands of methane, CH_4, which obscure the red region of the spectrum and are largely responsible for the greenish color of Neptune, the estimated amount of methane above the cloud level is about 2 5 km at standard temperature and pressure

Any amount of ammonia present would be frozen out of the atmosphere at the very low temperature of about 100°K that must prevail at this great distance from the Sun Theoretical considerations indicate that the bulk of the atmosphere must be hydrogen and helium, as is the case for the other three major planets; the presence of molecular hydrogen is directly confirmed by the observation of an absorption line at 8270 A, as is also true in the case of Uranus.

Satellites. Neptune has two known satellites. The larger and brighter, Triton, was discovered visually in 1846; the second, Nereid, which is very much smaller and fainter, was found photographically in 1949. Their main orbital elements are in the table.

	Mean dist to Neptune, 10³ miles	Sidereal period, days	Diameter, miles	Magnitude at mean opposition
I Triton	219 6	5 877	2500	(vis) 13 6
II Nereid	3730 ?	500	200	(phot) 19 5

The orbit of Triton is almost exactly circular and is tilted about 40° to the planet's orbit; the direction of the motion is retrograde Triton is too faint to show a measurable apparent diameter, and its linear diameter can be estimated only from its apparent brightness with an assumed value of the albedo

Triton's mass, however, can be directly determined from the periodic oscillations of Neptune about the center of gravity of the system; it is about twice the mass of the Moon. With this mass and the estimated diameter, the mean density is about 4, a higher value is hardly plausible and consequently the albedo can hardly exceed the assumed value of 0 2 Triton is therefore more similar to the brighter satellites of Jupiter than to those of Saturn and Uranus. It may even have traces of a methane atmosphere [G D V.]

Bibliography· H. N. Russell, R. S Dugan, and J Q Stewart, *Astronomy*, rev. ed., vol. 1, 1945.

Neptunium

Element number 93, neptunium, Np, a member of the actinide series of elements, was first synthesized by E M McMillan and P Abelson in 1940 by bombardment of uranium with neutrons to produce neptunium-239 with a half-life of 2 3 days. There are several isotopes of this element, but the one of mass 237, a long-lived α-emitter with half-life $2\,20 \times 10^6$ years, is the most important from the point of view of its chemistry.



The chemistry of its compounds, and consequently the oxidation states, was first elucidated by microchemical methods developed at the Metallurgical Laboratory of the University of Chicago during 1942–1945 These methods in conjunction with the x ray diffraction studies on anhydrous

Survey of the anhydrous compounds of uranium, neptunium, and plutonium

Coordinating anion	Uranium	Neptunium	Plutonium
O^{--}	UO_3, U_3O_8, UO_2	Np_3O_8, NpO_2	Plutonyl salts, $NaPuO_2Ac_3$, PuO_2, Pu_2O_3
F^-	UF_6, UF_5, UF_4	NpF_6, (NpF_5), NpF_4	PuF_6, PuF_5 ?, PuF_4, PuF_3
Cl^-	UCl_5, UCl_6, UCl_4	$NpCl_4$, $NpCl_3$	$PuCl_3$
Br^-	UBr_4	$NpBr_4$, $NpBr_3$	$PuBr_3$
I^-	UI_4	NpI_3	PuI_3
Oxyhalides	UO_2Cl_2, $UOCl_2$	$NpOCl_2$, $NpOCl$?, $NpOBr$?, $NpOI$?	$PuOCl_2$?, $PuOCl$, $PuOBr$, $PuOI$
S^{--}	US_2, U_2S_3, UOS	NpS_2 ?, Np_2S_3, $NpOS$, Np_2O_2S ?	Pu_2S_3, Pu_2O_2S

solid neptunium compounds by W. H. Zachariasen provided the means of identification of most of the neptunium compounds now known.

The chemistry of neptunium may be said to be intermediate between that of uranium and plutonium. These three elements form an interesting series in which to study trends in the stability of anhydrous compounds of the different oxidation states as a function of the electropositive (U, Np, Pu) and electronegative (O, F, Cl, Br, I, S) components of the compounds The small anions O^{--} and F^-, which polarize with difficulty, in general form compounds of higher oxidation number than do the more polarizable and oxidizable anions Cl^-, Br^-, S^{--}, and I^-. In aqueous solution, the positive ions of the different oxidation states are stabilized by solution and complex formation. In general it is easier to obtain the higher oxidation states of these elements in aqueous solution than in the anhydrous state.

In the table an attempt has been made to summarize available information with respect to the existence and stability of compounds of uranium, neptunium, and plutonium and to make estimates of relative stabilities The question marks signify those cases which, from observed trends, seem to be borderline. See ACTINIDE ELEMENTS; NUCLEAR CHEMISTRY; NUCLEAR REACTION; PLUTONIUM; RADIOACTIVITY; TRANSURANIUM ELEMENTS. [S.F.]

Bibliography: J. J. Katz and G T Seaborg, *The Chemistry of the Actinide Elements*, 1958

Nerve

A bundle of nerve fibers, or processes, passing to and from the brain and spinal cord and the body tissues. Functionally, motor nerves supply impulses to peripheral structures, such as muscles and glands Sensory nerves carry impulses from peripheral sense organs to the brain and cord. Both motor and sensory fibers are commonly present in one nerve In man, 12 pairs of cranial nerves and 31 pairs of spinal nerves pass from the brain and cord, respectively, to definite body segments. The cranial nerves emerge from the brain, the spinal nerves at successive vertebral levels. Both sets of nerves, with their branches, form the peripheral nervous system. See NERVOUS SYSTEM. [E.C.ST.]

Nervous system

A coordinating and integrating system which functions in the adaptation of an organism to its environment. An environmental stimulus causes a response in an organism when specialized structures, receptors, are excited. Excitations are conducted by nerves to effectors which act to adapt the organism to the changed conditions of the environment. In animals, humoral correlation is controlled by the activities of the endocrine system. This article considers the embryology, histology, and morphology of the nervous system, including the brain and cranial nerves.

COMPARATIVE EMBRYOLOGY

The complicated and varying anatomy of the adult nervous system in different vertebrates makes comparative embryological studies of these structures almost necessary for a sound understanding of their morphology Few fields in experimental analytical embryology have proved so fruitful as that of neural development A thorough study of

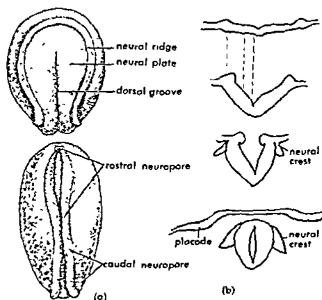


Fig. 1. Neural plate and its transformation to a neural tube in amphibians (a) Dorsal view. (b) Transverse sections.

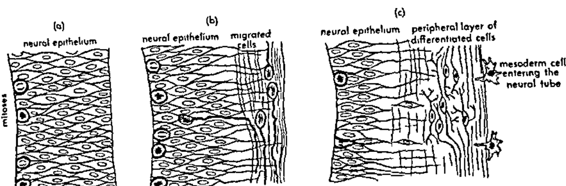


Fig 2 Schemes showing cellular constitution of the neural tube in three different stages (a-c) Ventricular wall is to the left in all figures

the embryology of the structures under study in animals used for experiments is necessary for a causal analysis. The embryology may be divided into a gross morphogenetic part, analyzing the development of the external and internal features of the nervous system; and a histogenetic part, dealing with the differentiation of the cells of the nervous system and their arrangement into nuclei.

Formation of neural plate and tube. The anlage of the nervous system is formed in the outer germ layer, the ectoderm, although some later contributions are also obtained from the middle germ layer, the mesoderm In most vertebrates a neural plate is formed, which later folds into a neural groove, then closes to form a neural tube. In some vertebrate species, such as the lamprey and bony fishes,

derm is due to inductive influences from underlying chordomesodermal structures *See* EMBRYONIC INDUCTION, EMBRYOLOGY, EXPERIMENTAL.

The closing of the neural tube in most species starts in the middle part of the embryonic body, the future neck region, and continues in a rostral and caudal direction Transitorily a rostral and a caudal neuropore exist

At the transition between the neural plate and the ectoderm a thickening, the neural crest, is formed (Fig 1) In the trunk, the ganglia of the spinal nerves are formed from it In the head some cells from the neural crest enter mesodermal structures, and others take part in the formation of cranial ganglia The latter are also formed from ectodermal thickenings lying further laterally, the so-called placodes *See* NEURAL CREST.

Histologic differentiation. At the site of formation of the neural plate in the ectoderm, the ectodermal cells elongate and form a cylindrical epithelium, the neural epithelium These cells continue mitotic division and form the primary germinal layer of the central nervous system. At a later stage of development (postneuromeric stage) cells migrate from the epithelium and form a peripheral layer (Fig 2).

Within the peripheral layer, and sometimes already within the neural epithelium, the differentiation of the cells proceeds towards neurons and glia cells via neuroblasts and spongioblasts, respectively. Also within the ganglia, formed from the neural crest and the placodes, a similar process of differentiation occurs.

From the surrounding mesenchyme (mesoderm) cells enter the central nervous system and form microglia cells These cells divide mitotically within the brain substance even in older embryonic stages Mitotic division of true neural cells outside



Fig 3 Microphotograph of a section through a young chick embryo, showing neuromeres as bulges of the nervous system.

the neural epithelium is probably of a low frequency. In the cerebellum, however, there is a thick proliferating layer of neural epithelial cells which also proliferates peripherally, the so-called embryonic granular layer.

After the formation of neurons and glia cells, the fibers which form give rise to the neuropile of the central nervous system, and to intra- and extracentral nerve bundles. The fibers emanating from the neurons of the ganglia grow as peripheral sensory fibers in peripheral nerves.

Morphogenesis. This aspect includes a consideration of the formation of neuromeres, the longitudinal structuring of the brain and cell migration. *See* ANIMAL MORPHOGENESIS.

Neuromeres. When the neural tube is developing, a segmentation of the central nervous system occurs by the formation of transverse bulges, neuromeres (Fig. 3). They are most distinctly seen in the hindbrain region, but can usually be identified in suitable embryonic stages in all parts of the central nervous system. They are most easily seen in vertebrate brains having a thin wall, for instance, those of sharks, birds, reptiles, and mammals.

Three different sets of neuromeric bulges develop successively, called proneuromeres, neuromeres, and postneuromeres. They represent a primary, secondary, and tertiary segmentation, respectively. The basis of neuromerism is the presence of proliferative patterns Each set of bulges thus corresponds to one period of increased proliferative activity in the neural epithelium, due to stimulative influences from underlying mesodermal structures. The proneuromeric segmentation extends from the neural tube into the neural crest and causes this to divide in the head region into portions, each corresponding to a proneuromere. This condition results in a topographic correspondence between the cranial ganglia and the neuromeres.

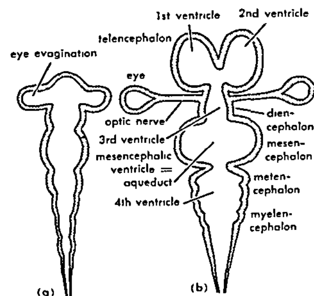


Fig 4. *Schematic horizontal sections through a vertebrate brain showing the transformation of (a) a neuromeric stage into (b) a brain vesicle stage.*

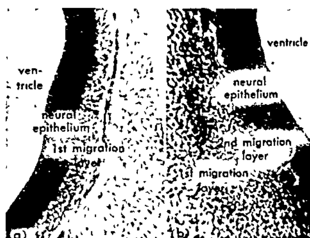


Fig 5. Two sections through the chick embryonic hemisphere, showing the neural epithelium and cells migrating from it. (a) One thin migration layer exists (b) *Two layers can be seen lying outside each other.*

At the time of neuromeric segmentation, the brain is subdivided into the so-called brain vesicles by local widenings of its lumen. In the rostral end more or less well-developed hemispheres are formed, in the middle of the brain anlage the mesencephalic bulge develops, and behind the latter the walls of the tube thicken into cerebellar folds. In this way the brain anlage is divided into five sections: the telencephalon, diencephalon, mesencephalon, metencephalon, and myelencephalon, and its cavity is divided into the rudiments of the adult ventricles (Fig. 4). The brain vesicles make the segmental characters of the neuromeric bulges less conspicuous.

When the postneuromeres develop, bulges can be identified only in the brain, not in the spinal cord. The presence of the postneuromeres influences the early development of the internal structures, giving them a slight segmental character.

Longitudinal columns. When the postneuromeric phase is at its height, a longitudinal structuring of the brain wall develops, consisting of four longitudinal bands of high proliferative and migrative activity. In the hindbrain these four columns approximately correspond to the anlagen of the four columns of functionally different qualities in the adult brain In the spinal cord the two dorsal columns fuse into one and the two ventral columns into another Rostrally, in the brain, the ventralmost columns stop at the rostral end of the mesencephalon and the dorsalmost column at the transition zone between the mesencephalon and the metencephalon (the isthmic region) The two middle columns build up the rest of the brain, and the borderline between them ends at the optic chiasma An approximate borderline between the two middle columns in the spinal cord and the myelencephalon is found in a furrow, the so-called limiting furrow of His It cannot be identified with certainty in the rostral part of the brain

The postneuromery and the longitudinal banding will give rise to a checkered pattern of prolifera-

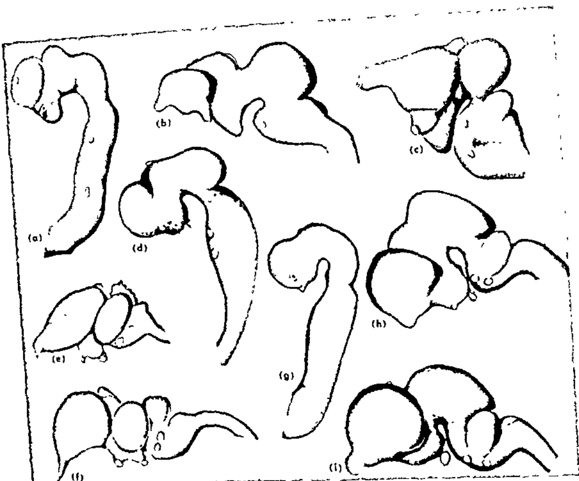


Fig. 6 Lateral views of embryonic brains of (a-c) a reptile (*Chelydra*), (d-f) a bird (*Melopsittacus*), (g-i) a mammal (*Spermophilus*) (After K. H. Krabbe)

tion centers. In the rostral part of the brain the transverse pattern will dominate, the longitudinal one in the caudal part.

Cell migration. Cell migration takes place from the neural epithelium into the peripheral or mantle layer. The presence of transverse and longitudinal proliferation centers will give rise to certain areas which are rich in cells, and cause a vivid lateral migration. Such areas are called migration areas, and their topography will be determined by post neuromery and longitudinal banding. The number and topography of the migration areas will be very similar in all vertebrate species.

The cells which have migrated laterally, may still lie in close contact with the neural epithelium and the ventricular wall, as in amphibians, or may lose contact with the epithelium and lie as a peripheral layer (Fig. 5). In many species, especially higher vertebrates, successive migrations of cells occur giving rise to two or more layers of such cells, situated concentrically. This feature is especially well marked in the hemispheres.

The migration layers may fuse or further subdivide into cell clusters, which represent the anlagen of the future brain nuclei. Therefore, they furnish the basis for comparative studies and homologizations of brain nuclei of different vertebrates.

Whole cell groups or brain nuclei may migrate (group migration), and in this way the topography of the nuclei may shift even from one brain vesicle to another.

Brain. In spite of the extraordinary variation in adult morphology of the vertebrate brain in different species, the early phases of development are essentially similar. The brain vesicle stages of a reptile, bird, and mammal are much alike, but owing to varying growth rates of different parts and to specialization processes the different patterns of the adult brains are formed (Fig. 6).

In a comparison of the embryology of the brains of anamniotes and those of [illegible]

[illegible] opment. In a shark or amphibian embryo the only marked bending is the cephalic flexure, situated in the same plane as the mesencephalon. This is also the first to develop in amniotes. In these brains, however, a nuchal flexure is also formed at the transition between brain and spinal cord anlagen, and later a pontine flexure ventral to the cerebellar

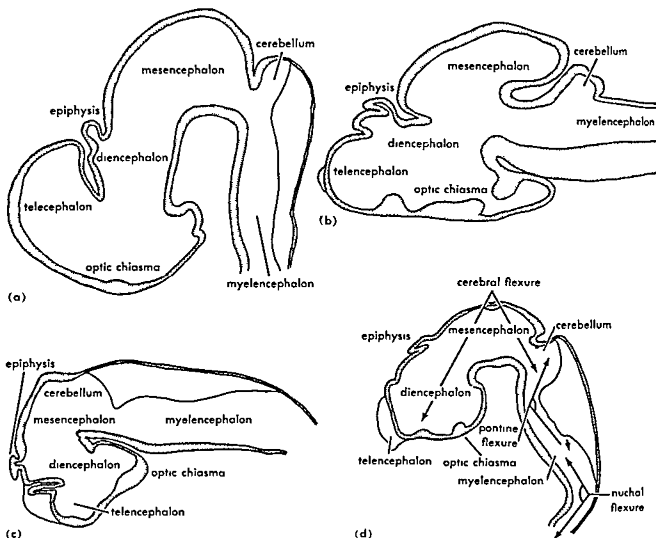


Fig. 7. Median sections of the brains of embryos of (*a*) shark, (*b*) bony fish, (*c*) salamander, (*d*) reptile

region (Fig 7). The flexures are most easily seen in median sections.

Telencephalon. The morphogenesis of the telencephalon of most vertebrates occurs by a lateral evagination or outbulging of the wall, giving rise to two hemispheric vesicles. In bony fishes, ganoids, and holocephalians, however, the lateral evagination is only faintly marked. Instead, a lateral bending, or eversion, occurs. The topography of the internal structures in the two kinds of forebrain will therefore be different (Fig. 8).

In all vertebrates two migration areas develop in the telencephalon—a dorsal one representing the

the different mantle regions and the septal and striatal nuclei develop.

Diencephalon. The morphogenesis of the diencephalon varies little in different species. A more or less well-developed transverse velum is formed *in the roof, caudal to which* the epiphyseal rudiment is situated. The paraphysis, which is included in the telencephalon, lies rostral to it. Part of the hypophysis develops from the bottom of the diencephalon while from the ventrolateral parts of the diencephalon the eyes are formed. The lateral walls are divided into a dorsal thalamic and a ventral hypothalamic region, containing the mammillary bodies. The hypothalamic region grows more in size in lower vertebrates than the thalamic does, while in higher vertebrates the opposite condition exists.

Mesencephalon The original single mesencephalic vesicle is divided into two vesicles which communicate broadly with each other In lower vertebrates this condition remains unchanged. The original wide ventricular cavity in higher forms is reduced to form the mammalian Sylvian aqueduct. The evaginations are connected ventrally with an unevaginated part, the tegmentum. Within the latter, the tegmental nuclei, oculomotor nuclei, and the red and black nuclei (nucleus ruber and nucleus niger) develop. The mesencephalic evaginations form the bigeminal bodies in lower forms and the quadrigeminal bodies in higher forms

Metencephalon. The cerebellum is formed in the dorsal part of the metencephalon. Its degree of development in different vertebrates varies considerably. The original raised lateral walls of the brain fuse to form a single plate. This is extremely compact, for example, in the bony fishes, and from its rostral end a so-called valvula grows rostrally. In *Petromyzon*, amphibians, and most reptiles the

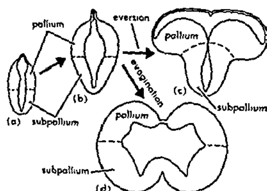


Fig 8 Schemes showing transverse sections of forebrains (a) *Primitive stage, develops via* (b) to (c) eversion or (d) evagination.

cerebellum remains as a simple transverse plate In sharks, birds, and mammals it develops into a dome, which may be more or less folded, thereby increasing its surface.

In all vertebrates a secondary proliferation layer, the embryonic granular layer, is formed in the periphery of the cerebellum; this layer disappears during later development The superficial layers of the cerebellum give rise to the mantle layer with the deep cerebellar nuclei being formed from deeper layers.

Myelencephalon This brain part remains in a relatively primitive state. Its roof is extended as a thin tela, its walls form a more or less V-shaped structure with only small variations The internal structures are dominated in their development by the above mentioned longitudinal columns

Spinal cord. The spinal cord remains as a comparatively slightly differentiated tube The primary lumen is secondarily reduced by the fusion of the side walls into a narrow central canal In the lateral walls, the longitudinal columns, separated by the limiting furrow of His, develop into the dorsal and ventral horns respectively In fishes the diameter of the spinal cord tube gradually diminishes in a rostrocaudal direction, but in four-footed animals intumescents develop level with the extremities by a process of degeneration of the regions situated in between.

PERIPHERAL NERVOUS SYSTEM

Cranial nerves. The cranial nerves are of quite varied morphological significance which is evident from their embryology

Olfactory nerve Fibers of the olfactory nerve grow out from the primary sensory cells of the epithelium of the nasal sac They run to the lateral surface of the telencephalic rudiment, usually entering it on the border between the pallial and subpallial regions

Optic nerve. The eyes develop as evaginations from the lateral walls of the diencephalon The stalks of the evaginations are the pathways of the future optic nerves. The neurites growing in from the retina within the stalk are thus really comparable to an intra-central brain fascicle They reach the brain in the floor of the diencephalon to form the optic chiasma. *See* Eye.

Ventral motor nerves. Neurites emerge from cells situated in the ventralmost longitudinal column of the brain stem and leave the brain surface as motor nerves. These nerves are the oculomotor, trochlear, abducens, and hypoglossal. The trochlear nerve fibers first grow dorsad, cross in the roof of the brain, and leave it dorsally in the fold between the mesencephalon and the cerebellum. The other nerves leave the brain ventrally. In *Petromyzon* the trochlear nerve nucleus develops dorsally at the site of the future crossing

Dorsal nerves The dorsal nerves (trigeminal, facial, statoacoustic, glossopharyngeal, and vagus) are all mixed nerves except the statoacoustic. The sensory fibers grow out from neurons differentiated within the cranial ganglia The motor fibers come from cells lying within the brain stem.

The cranial ganglia are formed from the head portions of the neural crest and from the ectodermal placodes. The neural crest is divided into four or five segments, called the thalamic (present only in lower vertebrates), mesencephalic, trigeminal, facial, and glossopharyngeal-vagus crests.

The placodes of lower vertebrates are made up of two groups, those associated with the lateral-line nerve system and called the dorsolateral placodes, and those situated further ventrally and giving rise to the main ganglia, the ventral placodes A summary of the vertebrate placodes is given below

	Placode	*Ganglion*
Dorso-lateral	Prelabyrinthic	Trigeminal, Facial
	Labyrinthic	Statoacoustic
	Postlabyrinthic	Glossopharyngeal, Vagus
Ventral	Ophthalmic	Trigeminal
	Epibranchial	Glossopharyngeal, Vagus

Spinal nerves. The spinal ganglia are formed from the neural crest which grows out like a continuous sheet from the dorsal margin of the neural tube and is secondarily split up into cell groups, the ganglia, by a segmentating influence from the somites Fibers grow out from the ganglionic cells and form the sensory fibers of the spinal nerves. Motor nerve fibers emerge from cells situated in the ventral horns of the spinal cord The ventral motor fibers and the dorsal sensory fibers fuse to form a common stem, which is again laterally divided into branches, innervating the corresponding segment of the body

Autonomic nervous system. The ganglia of the sympathetic nervous system develop ventrolateral to the spinal cord as neural crest derivatives. At first a continual column of sympathetic nerve cells is formed; it later subdivides into segmental ganglia The nerve fibers developing from these cells form the gray communicants to the

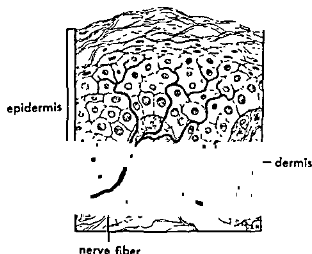


Fig 9. Schematic representation of a sensory nerve fiber passing through the dermis and terminating in free nerve endings among epithelial cells in the epidermis of the skin.

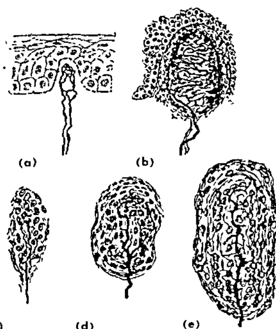


Fig. 10 (a) Nerve loops developing Meissner corpuscle from 7-month human fetus. (b) Adult Meissner corpuscle. (c-e) Stages in development of a Pacinian corpuscle in human fetuses at different approximate ages. (c) 3 months, (d) 3½ months, (e) 4 months (*From B. M Patten, Human Embryology, 2d ed., Blakiston–McGraw-Hill, 1953*)

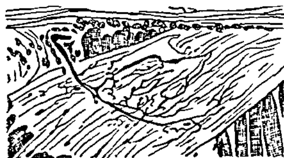


Fig. 11. Developing neurotendinous fibers from human fetus of 6 months. (*From B M Patten, Human Embryology, 2d Blakiston–McGraw-Hill, 1953*)

the peripheral sympathetic nerves. The white communicants develop from spinal cord cells. Along the peripheral nerve fibers, cells migrate to form the secondary plexi and ganglia.

The parasympathetic system is made up of preganglionic fibers, emanating as general visceromotor fibers from the brain and from the sacral cord segments. Cells migrate to form the peripheral ganglia along them. [B K.]

SENSE ORGAN EMBRYOLOGY

Sense organs. Groups of ganglion cells, connected with the brain and spinal cord, send tiny nerve fibers through cablelike nerves to various parts of the body where they pick up many kinds of sensations which keep the living organism in touch with its environment. Therefore, specialized receptor cells and nerve endings must be provided, especially over wide areas for such senses as touch, pressure, pain, temperature, and muscle and tendon sense. Wherever possible the description of the development of the special senses in the vertebrates is illustrated with human material. *See* SENSE ORGAN.

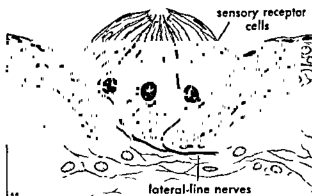


Fig 12 Schematic drawing of a lateral-line sense organ in the skin of an adult salamander, the common aquatic vermillion spotted newt Sensory receptor cells, surrounded by supporting cells and skin epithelium, terminate in hairs projecting into skin pores. Lateral-line nerves terminate around the bases of the sensory cells

Free nerve endings. Free nerve endings for pain and touch reach the skin as early as the third month in human fetuses. Their terminal branches then increase as the skin rapidly develops hair and nails from the fourth to sixth months (Fig 9). During this time certain terminal nerve fibers slowly become encased with specialized layers of flat cells. Some near the skin, the Meissner corpuscles, receive tactile stimuli. Nerve loops (Fig. 10*a*) near the skin gradually become encapsulated with specialized connective tissue cells (Fig. 10*b*). Others, Pacinian corpuscles (Fig. 10*e*), receive deep pressure sense and consist of more elaborate concentric layers of cells, like the sheaths of an onion, wrapped around a tiny nerve fiber. They develop in much the same way as the tactile organs (Fig. 10*c* and *d*).

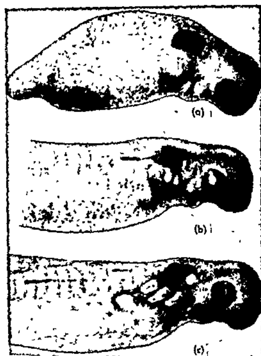


Fig 13 (a) Camera lucida drawing of a living salamander embryo of *Amblystoma punctatum* made 1 day after a lateral line placode was excised and replaced by a similar one (shaded) taken from a Nile-blue stained donor (b) Same living specimen as in (a), 24 hours later, showing one midbody lateral-line primordium migrating in the surface ectoderm toward the tail. (c) Same specimen as in (a) and (b) but 24 hours later than (b), showing a long and a shorter lateral-line primordium migrating down the side of the body and depositing clusters of blue-stained cells which form the sense organs ×10 (From L S Stone, The development of lateral line sense organs in amphibians observed in living and vital-stained preparations, J. Comp Neurol., 57(3) 507–540, 1933)

Continuing from the third fetal month many sensory nerve fibers spread over the body among developing muscle and tendon fibers, and as they branch, tiny flat plates develop at each nerve ending (Fig 11) A delicate fibrous network of connective tissue finally covers them The stimuli they pick up and relay to the central nervous system give the awareness of the position of the body and its parts

Lateral line organs. Some organs of special sense, such as the eye, are extremely complicated, whereas others are relatively simple In some aquatic vertebrates (many fishes and amphibians) there are lateral-line skin organs (Fig. 12) on the head and body, innervated by nerve trunks coming from cranial ganglion cells connected with the brain These organs apparently acquaint the animal with pressure changes in the surrounding water giving it a sense of orientation while swimming in a current, or a warning of an approaching object. There are no homologs in man or other animals These tiny pear shaped organs possess centrally placed, club-shaped sensory cells (Fig 12), each of which ends in a hairlike process at the free surface These cells are interspersed and surrounded by tall, flat, overlapping, leaflike supporting cells. Fine nerve fibers from the lateral-line nerve (Fig. 12) branch among the sensory cells to receive their stimuli. The apex of the organ communicates with a microscopic pore at the skin surface in amphibians, and with a canal system in the skin of fishes.

In amphibian embryos where they have been extensively studied, ectodermal thickenings, called placodes, first appear on the side of the head. Any one of these placodes can be stained with a blue vital dye (Fig. 13*a*) and as the embryo grows one can follow them as they elongate, migrate on the surface of the head and body (Fig. 13*b* and *c*), and deposit at regular intervals clusters of blue cells that form the lateral line organs. By this method one can observe the developing organs under the microscope as the blue dye particles migrate to the tips of the sensory and supporting cells (Fig 14) Each of them becomes innervated by the lateral-line nerve that follows the migrating placode This nerve comes from ganglion cells which are also placodal in origin A cluster of new secondary organs arises by a budding process from supporting cells of the primary organs (Fig. 15*a* and *b*). In practically all frogs and toads the lateral-line system degenerates at metamorphosis.

Taste buds. There are special chemical receptors somewhat like a rosebud in shape, called taste buds (Fig 17*a*), or gustatory organs They are common to all classes of vertebrates and function in a watery environment. They are associated with parts in the oral cavity, especially on the fungiform and



Fig 14 Camera-lucida drawing of a differentiated living lateral-line organ surrounded by two large pigment cells in the skin of a young 16 5-mm salamander larva, 16 days after operation shown in Fig 13a. Blue dye particles were observed during development as they migrated to the tips of the central sensory and surrounding supporting cells ×315 (From L S Stone, The development of lateral-line sense organs in amphibians observed in living

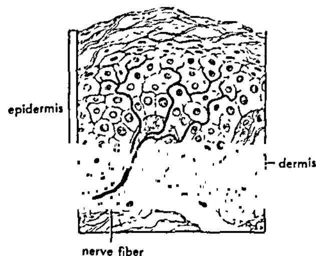


Fig 9. Schematic representation of a sensory nerve fiber passing through the dermis and terminating in free nerve endings among epithelial cells in the epidermis of the skin.

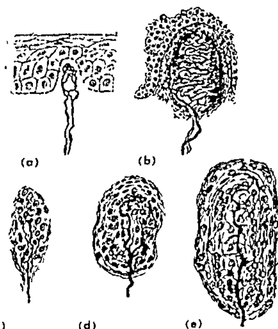


Fig 10 (a) Nerve loops developing Meissner corpuscle from 7-month human fetus. (b) Adult Meissner corpuscle. (c-e) Stages in development of a Pacinian corpuscle in human fetuses at different approximate ages. (c) 3 months, (d) 3½ months, (e) 4 months (*From B. M Patten, Human Embryology, 2d ed , Blakiston–McGraw-Hill, 1953*)



Fig. 11. Developing neurotendinous fibers from human fetus of 6 months (*From B. M Patten, Human Embryology, 2 Blakiston–McGraw-Hill, 1953*)

the peripheral sympathetic nerves. The white communicants develop from spinal cord cells. Along the peripheral nerve fibers, cells migrate to form the secondary plexi and ganglia.

The parasympathetic system is made up of preganglionic fibers, emanating as general visceromotor fibers from the brain and from the sacral cord segments. Cells migrate to form the peripheral ganglia along them. [B.K.]

SENSE ORGAN EMBRYOLOGY

Sense organs. Groups of ganglion cells, connected with the brain and spinal cord, send tiny nerve fibers through cablelike nerves to various parts of the body where they pick up many kinds of sensations which keep the living organism in touch with its environment. Therefore, specialized receptor cells and nerve endings must be provided, especially over wide areas for such senses as touch, pressure, pain, temperature, and muscle and tendon sense. Wherever possible the description of the development of the special senses in the vertebrates is illustrated with human material. *See* SENSE ORGAN.

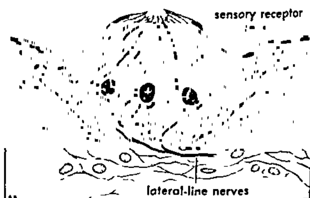


Fig. 12 Schematic drawing of a lateral-line sense organ in the skin of an adult salamander, the common aquatic vermillion spotted newt. Sensory receptor cells, surrounded by supporting cells and skin epithelium, terminate in hairs projecting into skin pores Lateral-line nerves terminate around the bases of the sensory cells

Free nerve endings. Free nerve endings for pain and touch reach the skin as early as the third month in human fetuses. Their terminal branches then increase as the skin rapidly develops hair and nails from the fourth to sixth months (Fig 9). During this time certain terminal nerve fibers slowly become encased with specialized layers of flat cells. Some near the skin, the Meissner corpuscles, receive tactile stimuli. Nerve loops (Fig 10*a*) near the skin gradually become encapsulated with specialized connective tissue cells (Fig 10*b*). Others, Pacinian corpuscles (Fig. 10*e*), receive deep pressure sense and consist of more elaborate concentric layers of cells, like the sheaths of an onion, wrapped around a tiny nerve fiber. They develop in much the same way as the tactile organs (Fig 10*c* and *d*).

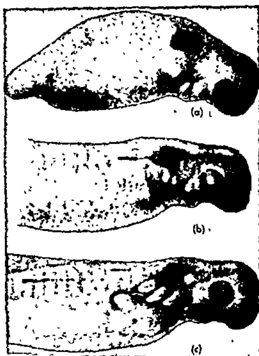


Fig 13 (a) Camera-lucida drawing of a living salamander embryo of *Amblystoma punctatum* made 1 day after a lateral-line placode was excised and replaced by a similar one (shaded) taken from a Nile-blue stained donor (b) Same living specimen as in (a), 24 hours later, showing one midbody lateral-line primordium migrating in the surface ectoderm toward the tail (c) Same specimen as in (a) and (b) but 24 hours later than (b), showing a long and a shorter lateral-line primordium migrating down the side of the body and depositing clusters of blue-stained cells which form the sense organs. ×10 (From L S. Stone, The development of lateral-line sense organs in amphibians observed in living and vital-stained preparations, J Comp Neurol, 57(3) 507–540, 1933)

Continuing from the third fetal month many sensory nerve fibers spread over the body among developing muscle and tendon fibers, and as they branch, tiny flat plates develop at each nerve ending (Fig 11) A delicate fibrous network of connective tissue finally covers them The stimuli they pick up and relay to the central nervous system give the awareness of the position of the body and its parts

Lateral-line organs. Some organs of special sense such as the eye, are extremely complicated, whereas others are relatively simple In some aquatic vertebrates (many fishes and amphibians) there are lateral-line skin organs (Fig. 12) on the head and body, innervated by nerve trunks coming from cranial ganglion cells connected with the brain These organs apparently acquaint the animal with pressure changes in the surrounding water giving it a sense of orientation while swimming in a current, or a warning of an approaching object. There are no homologs in man or other animals. [illegible] tiny pear-shaped organs possess several centrally placed, club-shaped sensory cells (Fig 12), each of which ends in a hairlike process at the free surface. These cells are interspersed and surrounded by tall, flat, overlapping, leaflike supporting cells. Fine nerve fibers from the lateral line nerve (Fig 12) branch among the sensory cells to receive their stimuli. The apex of the organ communicates with a microscopic pore at the skin surface in amphibians, and with a canal system in the skin of fishes.

In amphibian embryos where they have been extensively studied, ectodermal thickenings, called placodes, first appear on the side of the head. Any one of these placodes can be stained with a blue vital dye (Fig. 13*a*) and as the embryo grows one can follow them as they elongate, migrate on the surface of the head and body (Fig. 13*b* and *c*), and deposit at regular intervals clusters of blue cells that form the lateral-line organs. By this method one can observe the developing organs under the microscope as the blue dye particles migrate to the tips of the sensory and supporting cells (Fig. 14) Each of them becomes innervated by the lateral-line nerve that follows the migrating placode. This nerve comes from ganglion cells which are also placodal in origin A cluster of new secondary organs arises by a budding process from supporting cells of the primary organs (Fig. 15*a* and *b*) In practically all frogs and toads the lateral-line system degenerates at metamorphosis.

Taste buds. There are special chemical receptors somewhat like a rosebud in shape, called taste buds (Fig 17*a*), or gustatory organs. They are common to all classes of vertebrates and function in a watery environment They are associated with parts in the oral cavity, especially on the fungiform and



Fig 14. Camera-lucida drawing of a differentiated living lateral-line organ surrounded by two large pigment cells in the skin of a young 16 5-mm salamander larva, 16 days after operation shown in Fig 13*a*. Blue dye particles were observed during development as they migrated to the tips of the central sensory and surrounding supporting cells ×315. (From L. S Stone, The development of lateral-line sense organs in amphibians observed in living and vital-stained preparations, J Comp Neurol, [illegible]

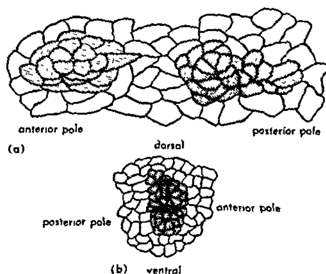


Fig. 15. Camera-lucida drawings outlining lateral-line organs (stippled cells) in process of budding. (a) Budding organ at posterior pole was derived 24 hours earlier from organ at anterior pole in the tail of larva (*Amblystoma punctatum*). (b) The lateral-line organ shown budding dorsally was observed to be derived by budding 24 hours earlier from the one ventral to it. None were observed budding anteriorly or posteriorly ×100. (*From L. S. Stone, The development of lateral-line sense organs in amphibians observed in living and vital-stained preparation, J. Comp. Neurol., 57(3):507–540, 1933*)

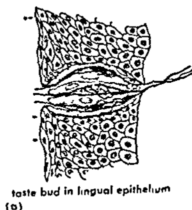
circumvalate papillae in the mammalian tongue, but in some fishes, such as the catfish (*Ameiurus*), many taste buds are also found in the skin, on the surface of the head and body. The central, rod-shaped sensory cells (Fig. 16*a*), neuromasts, are embraced by slender, overlapping, flat, supporting cells, the outer ends of which surround a pitlike excavation connected through a pore with the mucous epithelium of the mouth. These neuromasts, which send hairlike processes into the pit, are in contact with a basketlike network of nerve fibers They pick up the stimuli that are then carried by the nerve fibers to the cranial gustatory ganglia and on into the brain. *See* TONGUE.

In the tongue of the human fetus the taste buds first appear as clusters of epithelial cells and increase in number as the gustatory nerve fibers reach the epithelium. Although the taste organs are known to degenerate eventually after gustatory nerves are cut, their arrival at the epithelium in the first place may not be the stimulus which induces the organs to form. *See* EMBRYONIC INDUCTION.

It has been conclusively shown by experiments on salamander embryos that the lining of the floor of the future mouth can be transplanted from one embryo to the side of the body of another embryo; a tongue develops later with taste organs without having been innervated. It was also found that if the epibranchial ectodermal placodes on the sides of the head, which give rise to the gustatory ganglia, are excised, the taste organs develop normally without a nerve supply. How these special sense organs arise in any vertebrate is not known. Taste organs, like lateral-line organs, were found to increase in number by a continuous budding process from the peripheral supporting cells of older taste organs. In many vertebrates there is a continuous increase in taste buds for a long period. It is quite possible that this is accomplished by a similar budding process.

Very little is known about the time at which the taste organs become functional. Some investigators believe that significant reflex responses can be induced in premature 7-month infants by sweet, sour, and bitter tastes.

Olfactory structures. In man, the sense of smell also depends upon special neurosensory epithelial cells functioning in a moist environment within the nasal cavities The area of specialized olfactory epithelium lies in the upper deeper roof of the nasal mucous membrane and is made up of tall cells with bristlelike processes projecting into the mucus-covered surface where they act as chemical receptors (Fig. 16*b*) They are surrounded by tall supporting cells and extend towards the brain as thin fibers which contact fibers of intermediate ganglion or mitral cells. These in turn relay the olfac-



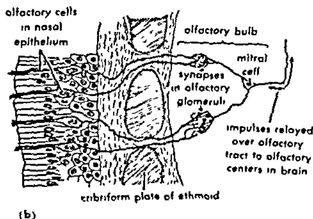


Fig. 16 (a) Taste organ from mammalian tongue. (b) Olfactory sensory cells of the mammalian nose related to the nerve tracts leading to the brain (*From B. M. Patten, Human Embryology, 2d ed., Blakiston–McGraw-Hill, 1953*)

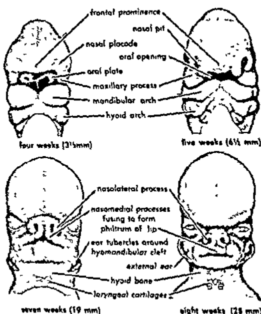


Fig 17 Drawings showing the development of the nose and other facial features in human embryos. (*From B M Patten, Human Embryology, 2d ed., Blakiston-McGraw-Hill, 1953*)

tory impulses along the olfactory tract to the appropriate centers in the brain.

The olfactory organs arise in a similar manner in all vertebrates, by an early appearance of a

... phibian embryos indicates that the formation of nasal placodes can be induced by neighboring mesoderm and brain-forming cells

In human embryos the nasal placodes appear during the fourth week (Fig 17*a*) Very soon the placodes sink inward, forming pits (Fig. 17*b*) which become deeper as the frame of the nose and

... The ... communicate shortly after the seventh week and ... the ... 17*d*) ... in the root of the two nasal passages, specialized sensory cells of the olfactory epithelium become surrounded by tall supporting cells (Fig 16*b*) Except for the skinlike lining at the entrance of the nares, all other areas of the nasal cavities become covered by columnar epithelium with surface cilia and mucus-secreting cells which keep the entire membrane covered with a moist film that provides the environment later for chemical stimulus of the hairlike ends of the sensory cells The rate at which full differentiation of this sensory mechanism takes place varies among the vertebrates The normality of the framework of the nose as well as the face and head depends a great deal upon the ability of the mesoderm to reach its full development. [L.S.ST.]

HISTOLOGY

Nervous tissue. One of the four primary tissues of the body is the nervous tissue, by which the functions of most other tissues are regulated, the environment perceived, the musculoskeletal system activated, and psychological processes generated. In vertebrates and in most invertebrates the nervous tissue is segregated into a central and a peripheral nervous system. The central nervous system is composed of groups of nerve cells and their incoming and outgoing nerve fiber connections. Each group or center is associated with a distinctive function. The peripheral nervous system in vertebrates consists of nerves, which are essentially cables of nerve fibers of diverse functions, and two types of nerve cell groups, or ganglia. The peripheral nerves contain fibers which originate in the central nervous system and conduct impulses to the muscles, as well as nerve fibers which originate in the ganglia.

Ganglia. The sensory ganglia which transmit nervous impulses directly to the central nervous system from the sense organs are considered as part of the peripheral nervous system. They are clusters of nerve cells within the course of the segmentally arranged nerve roots of the body, close to the entrance of the nerve roots into the central nervous ...

... gans, and transmit impulses to the central nervous system. This tissue in turn correlates sensory messages and relays impulses in coordinated fashion to the tissues, such as muscles, which perform appropriate responses.

The autonomic ganglia are composed of nerve cells which receive impulses from the central nervous system and relay secondary impulses to the glands, blood vessels, and smooth muscle fibers of the body wall and internal organs. Unlike the skeletal musculature, these effector structures are not under direct voluntary control, but respond to reflex stimulation and to strong emotional activation

Cellular components. The tissue of the central and peripheral nervous systems has important common structural and functional characteristics, as well as certain differences. In the central nervous system the basic pattern is that of elaborately structured neurons, communicating with each other in both simple and complex patterns and ...

Neuroglia The neuroglial matrix forms a supporting skeleton which extends from the central fluid-filled cavities or ventricles of the central nervous system to its outer membrane. This is covered by connective tissue similar to that found elsewhere in the body. The neuroglia cells are in large part intimately associated with the blood vessels and in-



Fig. 18. A nerve cell and surrounding neuroglia cells.

vest these as well as the nerve cells with their processes. Several types have been described: the astrocytes, oligodendroglia, and the ependymal cells which line the cavities or ventricles of the central nervous system. The so-called microglia, like the blood vessels, is derived from embryonic mesoderm, and invades the ectodermal neural tube in early embryonic development. The microglia cells become phagocytic during pathologic reactions of nervous tissue. A rich network of blood vessels constitutes the third important element of nervous tissues; this network invades the developing embryonic nervous tissue from the outer membranes, and is especially well developed in the nerve cell centers, or gray matter, whereas the white matter, which consists of bundles of nerve fibers like those of the peripheral nerves, is less well supplied with blood capillaries. *See* PHAGOCYTOSIS.

Nerve cells. Nerve cells are described as having a cell body and cytoplasmic extensions or processes. The nucleated portion of the nerve cell, as distinct from its protoplasmic extensions, is called the cell body (Fig. 19). Nerve cell bodies may vary in diameter from about 7 to over 70 μ. The single threadlike axon process, however, may range in size from a millimeter or less to many feet in length, as for example the axons of the motor nerve cells in the spinal cord which extend as far as the muscles of the foot

Nerve cells differ from other cells in several other respects. After reaching maturity before or soon after birth they never again undergo cell division. This means that nerve cells destroyed by disease can never be replaced, but it is this inability to reproduce which makes possible stable patterns of communications within the nerve fiber pathways of the nervous system.

The cytoplasm of the cell body is characterized by the presence of nucleoprotein aggregates of the pentose type, called Nissl bodies, or chromophilic substance. These are more conspicuous than in most other cells, and their stainability with basic aniline dyes permits a convenient identification of nerve cells and cell groups Moreover, the chromophilic substance is highly responsive to cell injury and is therefore a useful indicator of damage to the nerve cell by noxious agents such as viruses or toxic materials, by mechanical trauma, or by nutritional deficiencies. The pentose nucleoprotein of the cytoplasm of nerve cells is believed to play a role in the growth and maintenance of the axon process.

As distinct from the growth and maintenance functions of the nerve cell, the specific function of transmission of the nervous impulse is subserved by electrochemical processes which take place at the nerve cell membrane. This membrane, as seen by electron microscopy, is a specific differentiation of the nerve cell cytoplasm, and resembles the boundaries of other cell types

Within all parts of the cytoplasm of nerve cells are found delicate parallel arrays of fibrillar structures, the neurofibrils. These have been observed in living nerve cells, and have an affinity for silver salts which makes possible the selective staining of nerve cells with a variety of silver impregnation methods. The functional role of the neurofibrils is unknown.

Nerve fibers. The threadlike outgrowth of the cell body in embryonic life, the nerve fiber, or axon, remains dependent upon the cell body subsequently, in the sense that interruption of continuity of the axon results in death of the fiber beyond the point of interruption. In fibers of the peripheral nervous system, the portion of the fiber still connected to the cell body may, however, regenerate, and in favorable instances may reinnervate the structure to which it was previously connected, such as muscle or skin. Nerve fibers may be less than 1 μ or as large as 1 mm in diameter. The rate of conduction of nerve impulses varies in proportion to the diameter of the nerve fiber. The nervous impulse is a wave of excitation which is accompanied by changes in electrical potential which have been measured and studied in detail by means of the oscilloscope, a modern recording and amplifying device based on the development of the cathode ray tube The conducted excitation is believed to be the result of local permeability changes

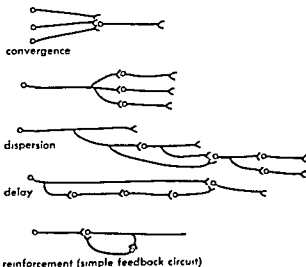


Fig 19. Diagram of certain patterns of nerve-cell circuits.

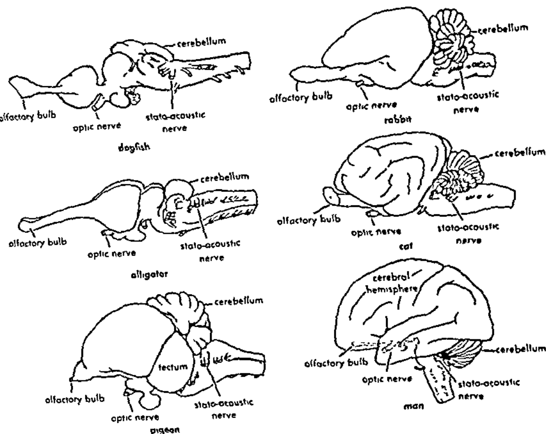


Fig 20 Lateral views of the brains of several vertebrates drawn approximately the same size, instead of to scale, to show relative development of various parts

in the axon membrane which produce differences in concentration of ions across the surface boundary

Neurolemma Nerve fibers are sheathed in a segmental fashion by rows of neuroglialike cells which in the peripheral nerves are known as cells of Schwann These cells form the neurolemma sheath. The larger fibers are also encased by a lipoprotein or myelin sheath which is produced by the Schwann cell and is therefore also segmentally arranged along the nerve fiber Recent studies with the electron microscope have clearly revealed a laminated structure of the myelin sheath, and have suggested that this is formed in embryonic development by a wrapping of the Schwann cell cytoplasm around the axon almost as one would wrap a bolt of cloth The ends of the myelin segments are not quite in contact with each other, and the interruptions of the insulating sheath, or nodes of Ranvier, are believed to play an important role in the conduction of the nervous impulse External to the myelin sheath is found a delicate sheath of connective tissue origin the endoneurial sheath. This is continuous with the connective tissue sheaths of bundles of nerve fibers (perineurium) and with the outermost connective tissue sheath of the nerve, the epineurium.

Dendrites The receptive processes of a nerve cell, the dendrites, are encased by neuroglia cells in the central nervous system, but when such processes extend into peripheral nerves, they too may be ensheathed by myelin, as are the axon processes. Such sensory dendrites extend into the skin, muscular tissue, and viscera, where they terminate as nonencapsulated sensory endings or as encapsulated endings, such as Merkel's corpuscles, Pacinian corpuscles, Meissner's corpuscles, muscle spindles, and the tendon organs of Golgi. In such endings the capsule is formed by cells related in origin to the Schwann cells of the nerve fiber.

Patterns and connections. Nerve fibers tend to be unbranched, but usually arborize widely before terminating. The termination of each arborization is usually knoblike as it makes contact with dendrites or cell bodies of other nerve cells, or with the membrane of muscle cells. Whereas nerve [illegible] pulses [illegible] a fibe [illegible] direction at the points of contact of one nerve cell with another The surfaces of contact or synaptic junctions also make possible the modulation of nervous transmission, because the properties of the receptive surfaces, the pattern of discharge of impulses among the synaptic endings on a nerve cell, and the physiological state of the nerve cell determine whether it will or will not become excited by incoming signals (Fig 20) Thus the firing of a nervous impulse by a nerve cell is determined

by synaptically transmitted signals arriving from several or many sources. The number of nerve cells, number and variety of sources of signals arriving within it, and the array of synaptic patterns within it determine the complexity and characteristics of its functional role. The end-to-end linked neuron chains in the following examples of neuron patterns can be seen to make possible the convergence, dispersion, delay, or reinforcement of nervous impulses passing from one relay center to another. Additional modulation of function is made possible by the fact that synaptic axon terminals may be related to dendrites, to nerve cell bodies, or even to the axons of other nerve cells. The latter relationship is less common than the others, but makes possible the bypassing of dendrite or cell body of the receptive nerve cell.

The nervous tissue is thus composed of nerve cells separated from each other by an equal or greater number of supporting neuroglia cells, except at special points of contact of nerve cell with nerve cell. Since the nerve cells in the human brain are numbered in the billions, and their relationships with each other are complex, an infinite variety of sensory and behavioral experience can be assumed to be represented within the organ of cerebration or sentience. At the present time, however, the differences between human experience and behavior and that of other mammals cannot be explained by observable differences in the basic structure of their respective nervous tissues, nor by brain size and numbers of nerve cells and their connections Accumulating evidence rather suggests that behavior differences among various species of animals may be the reflection of higher levels of organization of cerebral centers as related to each other and to sensory and effector organs of the body. See BIOPOTENTIALS AND ELECTROPHYSIOLOGY; REGENERATION (BIOLOGY); SENSATION [D.B.]

COMPARATIVE MORPHOLOGY

General principles of structure. The nervous system of man has essentially the same pattern as that of other vertebrates and the nervous system of this large group is in many fundamentals like that of lower forms, although the external shapes vary widely (Fig 20).

In all animals that are made of more than a few cells there is a specialized type termed the neuron, or nerve cell, which is characterized by having long processes capable of conducting stimuli. Such cells do not occur in the plant world or in the simplest animals. The entire nervous system of vertebrates is made of neurons and certain supporting elements plus numerous necessary blood vessels.

Neurons are primarily concerned with the transmission of impulses between different parts of the body as required in the coordination of the animal's activity. Through the nervous system an animal can receive stimuli (that is, information about changes in the environment, external or internal), by way of its various sensory receptors, and responds to them by excitation of effectors (voluntary, involuntary, and cardiac muscle and glands).

Structure and function of neurons. Neurons vary in size but most are comparatively large cells. They may be rounded or irregular, and are named by shape or by the number of processes they possess. Of processes there is usually one axon (sometimes two) and there may be one or more dendrites The cell body of a neuron, although its material substance may be changed in normal function, is remarkably designed to persist throughout the life of the individual. Once laid down in the final state of development in the fetus, neurons may grow and extend the length of their processes, but are not reproduced if destroyed.

Each neuron has a large clear nucleus and prominent nucleolus, both of which are intimately involved in the metabolism of the cell and especially in the synthesis of nucleoproteins.

The processes of nerve cells, axons, and dendrites have different structure; dendrites are mostly short (though at times elaborate in form) protoplasmic extensions of the cell body, whereas axons may run for long distances and are the true nerve fibers of central nervous pathways and peripheral nerves. The axons may be bare or covered with either one or both of two different tubular sheaths: the myelin sheath, a laminated lipid layer, and the neurilemma, a nucleated cellular tube. Only fibers in the peripheral nervous system possess a neurilemma

Synapse. Synapses between neurons occur in several morphological patterns, but essentially the connection is one of close contact between the ends of branching axons and the body, dendrites, or base of the axon of another neuron; the separation of the two neuronal membranes is about .02 μ. Sensory endings vary in form with the function of the receptor and the histologic type of axon involved. Certain of them are bare (as endings concerned with pain) and others are covered by specialized cells (as tactile endings) The most elaborate types of specialization occur in the eye and the ear with accessory structures to aid in the activation of the nerve fibers that finally conduct the impulses of sight and sound Endings of nerve fibers that supply motor impulses to muscles and glands vary with the tissue innervated See EAR, EYE, PHOTORECEPTION, VISION.

Reflex arc. The neuron is considered the structural unit of the nervous system, it is customary, however, to consider a simple reflex path (arc) as the functional unit because it involves a complete mechanism for response to stimuli The simplest form of reflex arc consists of (1) a primary sensory neuron with sensory endings, for example, in the skin, and (2) a primary motor neuron with endings on voluntary muscle, the two neurons are in communication in the central nervous system by contact at synapses. But most reflex paths include more than two neurons, there being one or more association neurons placed between the primary motor and sensory ones, and having synaptic contact. This arrangement allows spreading of the sensory impulses not only to more motor neurons but makes possible bringing into action higher levels of the nervous system where many in neu

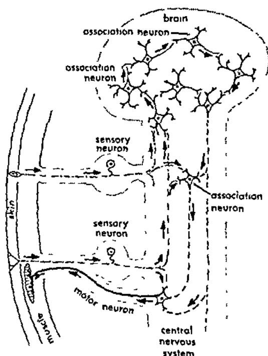


Fig 21 Diagram of general structure of nervous system Sensory neurons innervate the skin and send centrally directed fibers into the central nervous system to synapse with association neurons Through some of the latter, lower-level reflex connection is made to motor neurons

... neurons.

rons are available for the organization of more varied types of response to stimuli (Fig 21) In these higher levels the hereditary stamp of species behavior (instinctive) as well as the accumulated experience (memory) of the individual may cause modification of activity, reflex or otherwise. *See* BEHAVIOR AND HEREDITY, BEHAVIOR, ONTOGENY OF INSTINCTIVE BEHAVIOR

Information which is collected by various types of sensory endings exclusive of special senses can be described in familiar terms The external surface of the body is sensitive to pain, temperature changes, touch, and pressure, and these are usually well ...

less

the

of t

beir ... stretching or tension is an adequate stimulus for pain of visceral areas Localization of touch, pressure, or pain from viscera is relatively inaccurate Within muscles, joints, and tendons there are nerve endings called proprioceptive which give accurate information as to the position of a part. Such sensations enable one to touch the end of his nose, for example, while his eyes are closed. The proprioceptive sense is the necessary source of the sensory side of much activity of skeletal muscle in normal movements. Vision and hearing and equilibration also contribute constant controlling influences *See* EQUILIBRIUM, BIOLOGICAL; SENSATION, TAXIS

Central pathways. In the spinal cord and brain stem, nerve fibers carrying the various sensory modalities make numerous reflex connections and at the same time connect with long pathways which ascend to higher levels of the brain conveying more or less specific sensations. These long pathways have been late additions and are better developed in the higher vertebrates, but this does not diminish the basic functional pattern of activity represented by more diffuse local reflex connections, or more widespread connections such as occur in the reticular formation. The reticular formation can be considered as the great background of *central nervous system paths*, both ascending and descending, which are fundamental to all vertebrates, and out of which the special paths of higher forms are segregated. Among others, *for example*, a special pathway is provided for pain and temperature; *it* ascends along the ventrolateral part of the cord to the thalamus, the lateral spinothalamic path This is at times cut by the neurosurgeon to eliminate continuous and otherwise *uncontrollable* pain. Other paths in the dorsal aspect of the *cord* carry *sensations* of touch, pressure, proprioception, and the impulses necessary for tactile localization and discrimination (Fig. 22). These and the spinothalamic paths are relayed *through* the thalamus to sensory areas of the cerebral *cortex for* conscious *appreciation* of sensations. *Additional* paths carry similar types of impulses *to the cerebellum*, which is not involved with consciousness but with proper *fine adjustment* of the muscles during activity *See* RETICULAR FORMATION (BRAIN); SENSE ORGAN

In general the pattern of the sensory side of spinal nerves is duplicated by the cranial nerves with due allowance for specialization in function and anatomical arrangement By somewhat *similar* patterns all the sensory impulses destined for conscious appreciation eventually reach the level of the thalamus and then are relayed to the basal ganglia and to several different areas of the cerebral cortex. From these higher levels impulses descend to the *motor neurons* activating effectors that produce the responses. *See* CRANIAL NERVE; NEURON.

Cortical representation. A discussion of higher levels in the brain therefore involves the areas of arrival of the ascending impulses, the areas of departure of descending impulses, and the *intermediate areas of association* (for lack of a better term). These are not all sharply separated and in the brain of man certain *areas* have more representation than they do in lower vertebrates ... cortex is ...

tex a

recei

part ... the temporal lobe *on* the side of the ... cortex of the cerebrum receives auditory impulses *In* an area roughly superior to this, extending to the top of the head, *in* the parietal lobe *sensations* are received from the

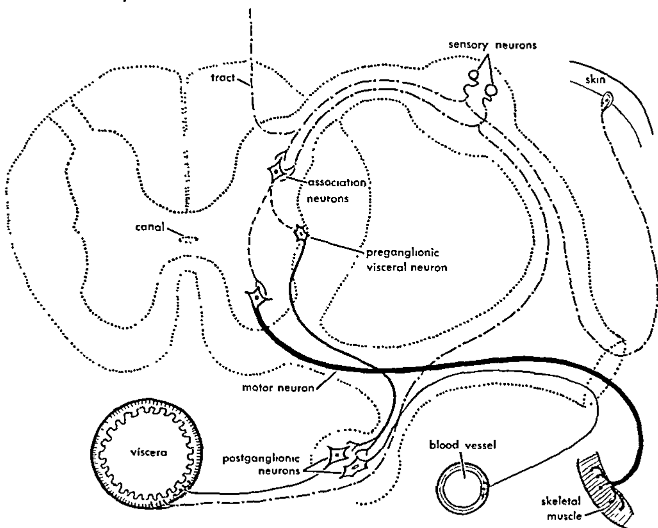


Fig 22. Diagram of spinal cord showing arrangement of neurons of spinal nerves Sensory neurons in a spinal ganglion send peripheral nerve fibers to the skin and to the viscera, and central branches to the spinal cord for synapse with association neurons and for ascending tract.

opposite side of the body (somesthetic area). Adjacent to these areas are zones in which the sensory experiences are correlated with other cerebral functions. *See* SOMESTHESIS.

Immediately anterior to the somesthetic area across the central sulcus is the motor area. Within this zone electrical excitation elicits movements of various parts of the body on the opposite side. Definite subdivisions corresponding to fingers or toes can be located, and it is of significance that in man's brain the hand and arm have larger representation than the foot and leg. In lower forms, as the cat and the goat, a difference between representation of extremities still is present but is not as marked in degree. With man's hand free from locomotion there is a great increase in the variety of motion of its parts. Parallel with this freedom of the hand man has acquired mechanisms for speech which set him apart from lower forms and has made possible the storage of experiences (through memory and written speech) and therefore the development of all the artifacts of civilization.

The speech mechanism is developed in the dominant hemisphere, that is, on the left side in right-handed people, although there is evidence that speech appears also partly represented on the nondominant side. *See* SPEECH.

Control of voluntary motion. From the motor area of the cerebral cortex and other regions of the cerebrum descending fibers pass to the brain stem and spinal cord to end synaptically on cells near, but probably not on, primary motor neurons (the final common path), which send their axons through cranial or spinal nerves to skeletal muscle. While direction of voluntary activity from cerebral cortical areas is conveyed over the corticospinal tract, there are other more indirect paths through large masses of nerve cells within the cerebrum, termed basal ganglia, which govern and steady movements. Interconnections of higher and lower levels with the cerebellum bring postural adjustments and fine synergy of action into control of voluntary muscle Furthermore, there are constant reflex influences from the vestibular apparatus of the internal ear, quick adjustments to visual and auditory stimuli, and reflexes at intersegmental spinal levels from external and proprioceptive stimuli to assist in the final action

Visceral mechanisms. The control of visceral mechanisms goes on in the body of man or lower animals in rather automatic fashion with less necessity for voluntary control than skeletal muscles, but distinctly regulated and affected by nervous mechanisms in all levels of the ner tem [illegible]

Fig 23 Schematic drawing of a vertebrate hemisphere in medial view with adjoining part of the diencephalon The hemisphere is cut into two parts by a cross section, and the diencephalon is sectioned medially. Section surfaces are cross striated

cerebral cortex to cord The motor outflow to visceral structures is termed for convenience the autonomic system It is generally slower in response to reflex stimuli than that of voluntary muscle, and visceral activity is often prolonged beyond the time of the nervous action by the production of humoral substances, for example, epinephrine, which spreads in the blood stream and maintains the influence on visceral effectors This is shown in a frightened animal by the dilated eyes, more rapid heart beat, and erection of the hair.

Certain cerebral cortical areas have been shown to have marked connections with visceral mechanisms through the hypothalamus, however, the latter has within it regulating mechanisms of great importance to the homeostasis of the internal environment In it are located nervous mechanisms necessary for the regulation of water intake and output, temperature control, metabolism of fat storage, and a variety of other functions, many of which are further regulated by hormones from glands of internal secretion The most influential of these glands, the hypophysis or pituitary, is in part an outgrowth of the hypothalamus See HOMEOSTASIS, THERMOREGULATION

Integrative action. The responses made to changes in environment are usually such as to be of use to the whole animal It is apparent that the responses to specific stimuli can thus be either stereotyped or varied It is also evident that the fewer the association neurons involved in a response to stimulus the more stereotyped the response is likely to be, and the greater the number the greater the possibility for variation of the response Man as the highest on the vertebrate scale, has proportionately the greatest development of association neurons, which form the bulk of his brain In man's capacity for choice, that is, for varying the response ...

ment

beha

ries

attempts at adaptation to external forces have resulted in the building up of an elaborate culture. Within this frame of reference the principles of natural selection and survival of the fittest are observed to be effective. See NERVOUS SYSTEM (INVERTEBRATE); NEUROPHYSIOLOGY; SPECIALIZED TISSUE [S.L.C.]

BRAIN

That part of the vertebrate central nervous system lying within the skull is the brain Compared with the other component of the central nervous system, the spinal cord, the brain is characterized by the enormous accumulation of nervous substance and its specialization in connection with the particular sense organs present in the head. The boundary between the brain and the spinal cord is fairly arbitrary, and has no sharp counterpart in the internal structure

Subdivision of the brain. For embryological reasons the vertebrate brain is subdivided into five sections, rostrally to caudally; these are the telencephalon, diencephalon, mesencephalon, metencephalon, and myelencephalon

Telencephalon In most vertebrates this region is represented by the two hemispheres, which in higher vertebrates make up the biggest part of the brain A brain ventricle is present within each hemisphere, communicating through canals, the foramina of Monro, with the rest of the brain ventricle system In association with the olfactory organs, the olfactory bulbs extend from the rostroventral part of the hemisphere, usually connected with the rest of the hemispheres by stalks, the olfactory tracts. The hemispheres are made up of dorsal parts, the mantle layer or pallium, and ventral parts, the striatal body laterally and the septum formation medially (Fig. 23)

In higher vertebrates the great expansion of the hemispheres results in the development of different lobes In the human brain a frontal, a parietal, an occipital, and a temporal lobe can be separated on each side The surface of the hemisphere may be smooth or may develop a system of convolutions, gyri, separated by furrows, sulci, In this way the surface of the hemisphere is considerably increased

Diencephalon Within the diencephalon the brain cavity, the third ventricle, is unpaired Its roof is made up of a thin membrane, the choroid tela.

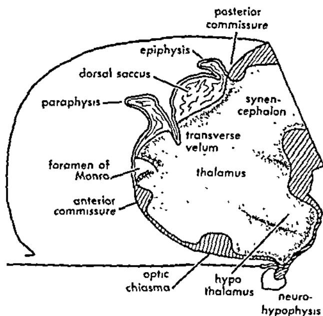


Fig. 24 Schematic drawing of vertebrate diencephalon in medial view. Section surface is cross striated. Contours of a hemisphere are indicated.

Some structures lie within the roof. Among these are the transverse velum, the dorsal saccus, the epiphysis, the paraphysis, and in some species one or two parietal eyes. In the floor of the third ventricle the optic nerves cross as a so-called chiasma. The neurohypophysis lies caudal to it in higher vertebrates. A saccus vasculosus, a richly vascularized thin-walled bag, lies in this position in fishes In the lateral walls the brain substance may be divided into a dorsal thalamic mass and a ventral hypothalamic mass (Fig. 24). Two pairs of geniculate bodies are situated laterally in higher forms. Between the true diencephalic cell masses and the mesencephalon, a synencephalic region lies interposed, which embryologically is to be derived from the diencephalon

Mesencephalon. The mesencephalic ventricle is wide in lower forms but is reduced to a narrow canal in mammals, the Sylvian aqueduct. The dorsal parts of this brain region form the so-called optic tectum in lower forms; in mammals the quadrigeminal bodies are formed here. A tegmentum is ventrally situated and from this region the trochlear and oculomotor nerves emerge.

Metencephalon. The dorsal part of the metencephalon is formed by the cerebellum. The ventral part is directly continuous with the myelencephalon in lower forms, but a pons region develops at this junction in mammals. The cerebellum consists of lateral auricles and a median corpus in lower forms; in higher vertebrates small lateral formations, the flocculonodular lobes, correspond to the auricles and carry lateral hemispheres. An unpaired vermis corresponds to the corpus (Fig 25).

Myelencephalon. This is also called the medulla oblongata. Its ventricle has a thin choroid tela as a roof. The caudal limit of this tela is a rough border against the spinal medulla. In dorsal view the ventricle of the myelencephalon, together with the metencephalon, has a rhomboid appearance, which is why this brain section is sometimes called the rhombencephalon. All brain nerves emerge from the myelencephalon except those already mentioned. The cell material in the lateral walls of the myelencephalon is arranged into cell columns, orientated longitudinally, each column having a single functional quality.

Function of the brain divisions. In lower vertebrates, such as fishes, the different brain sections are strongly dominated by the associated sense organs. The olfactory organ connects through the olfactory nerve to the olfactory bulb. New olfactory bundles extend from it and reach secondary olfactory centers in the telencephalon. This region is mainly an olfactory brain in lower vertebrates. In the same way the mesencephalon is dominated by the optic radiation, and the myelencephalon by the nerves connected with the branchial region and the lateral-line system. The efferent system of the head is mainly located in the myelencephalon also with the cerebellum as a correlating organ of locomotion. This basic pattern is retained in the brains of higher vertebrates, including mammals and man; however, a large and important shift in function takes place. The telencephalic structures lose their primary connection with the olfactory apparatus to a large extent, and develop into a superimposed correlating organ for more caudal levels of the brain The thalamic part of the diencephalon increases, being an important relay station between the increasing mass of telencephalic tissue and more caudal levels.

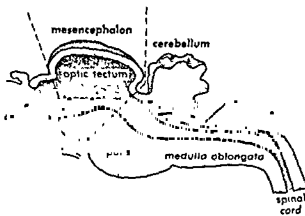


Fig 25 Schematic drawing of caudal part of a vertebrate brain in medial view Section surface is cross striated

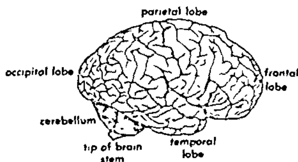


Fig 26 Drawing of human brain in ' ' view.

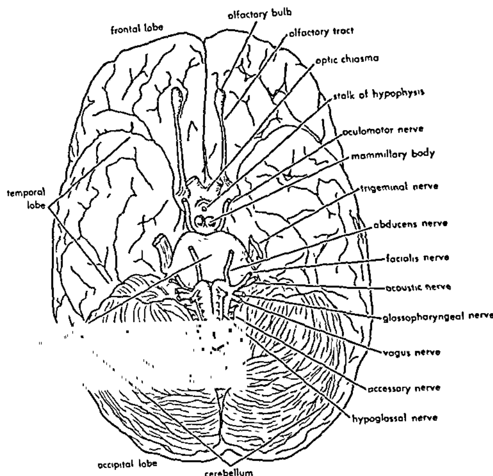


Fig 27 Drawing of human brain seen from below

In higher animals a sort of centralization process has thus occurred with an increase in size and significance of the telencephalic part, the brain stem which acts as a correlation center over the rest of the brain This mode of differentiation seems to be the basis for more complex intellectual activity

The particular degree of development of the specific sense organs, the varying capacity of locomotion and finally the above-mentioned gradual centralization process cause large modifications in size and degree of differentiation of the various parts of the brain in different species

Phylogenesis of the brain. The highest degree of telencephalic development is reached in mammals and man, resulting in an enormous degree of development of the cortical layers of the pallium Impulses from the spinal cord and the brain stem reach this part of the brain by means of fibers passing through the thalamus, which therefore also increases in size The fibers emerging from the cortex and descending to lower brain levels form a so-[illegible]

is present in reptiles also and in a rudimentary form in amphibians, and the latter already occurs in fishes. Similarly a neothalamus can be separated from a paleothalamus, the former being associated only with the neopallium, the latter having subcortical connections In the cerebellum a neocerebellum may be distinguished, having cortical connections via the pons, from a paleocerebellum, being connected with brain stem and spinal cord centers In a similar way the striatal body is subdivided into a paleostriatum, an archistriatum, and a neostriatum During ontogenesis, however, rudiments of all these parts can be found in nearly all vertebrate species It is thus not a question of adding new morphologic units during phylogeny, but of an increasing development of certain parts, together with a shift in functional characteristics.

Human brain. The weight of the adult human brain is about 1300 g, the greatest part of which is made up of the hemispheres. The system of sulci and gyri of the hemispheric surface is very well developed (Fig 26). It is possible to determine anatomical localizations for different functions The cerebellum is situated behind and beyond the hemispheres This, too, is grooved considerably and consists of a great number of gyri. Well-developed cerebellar hemispheres can be distinguished from the median vermis. Part of the brain stem can be seen at the base of the brain, that is, the medulla *oblongata* with its nerves, the thick commissural bundles that make up the pons, the bottom of the

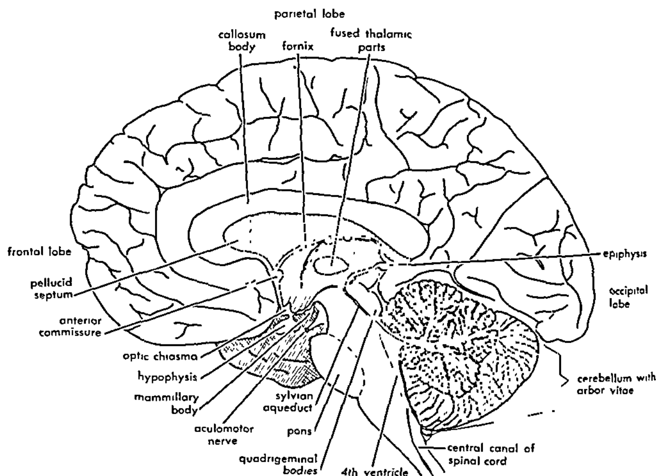


Fig. 28. Drawing of human brain, sectioned medially

diencephalon with the mammillary bodies, the *hypophysis, and the crossing of the optic nerves* On the ventral surface of the two hemispheres, the olfactory bulbs and tracts can be seen (Fig 27).

A peculiar distribution of white and gray matter of the cerebellum produces the so-called arbor vitae which can be seen in a median section of the brain (Fig 28). The roof of the mesencephalon appears as the quadrigeminal bodies. Between the two hemispheres lies an enormous commissural bundle, the corpus callosum Two columnar fiber bundles, the fornices, lie on each side below and behind it connected by a hippocampal commissure. The fornices describe a half-circle from the *temporal lobes upward, forward, and downward* to end in the mammillary bodies. They form the anterior limit of the foramina of Monro, which connect the ventricles of the hemispheres with that of the diencephalon. Between the corpus callosum and the fornices the brain wall is much extended to form a thin septum, the pellucid septum.

Meninges. The brain is enclosed and protected by three membranes of connective tissue. The outermost one is called the dura mater, and also represents the internal periosteum of the skull in the adult The innermost one is called the pia mater and is a thin membrane, lying close to the brain *surface and extending into the grooves on the brain* surface The arachnoid lies between these two membranes. The space between the arachnoid and the pia is filled by the cerebrospinal fluid, which further protects the brain This fluid is produced in *the brain ventricles* from the choroid plexi in the roofs of the third and fourth (myelencephalic) ventricles, fills the ventricular system, leaves it through openings in the myelencephalic tela, and flows into the subarachnoid *space, where it is absorbed.*

CRANIAL NERVES

The cranial or cerebral nerves are the peripheral *nerves of the head*, being related to the brain. The number and degree of development of the nerves varies in different species. The functional quality of the different nerves also varies Twelve pairs of cranial nerves have been *distinguished in human* anatomy, and these nerves have been numbered rostrally to caudally in the following way:

I–Olfactory nerve, fila olfactoria
II–Optic nerve, fasciculus opticus
III–Oculomotor nerve
IV–Trochlear nerve
V–Trigeminal nerve, in most vertebrates divided into three branches: ophthalmic, maxillary, and mandibular
VI–Abducens nerve
VII–Facial nerve
VIII–Statoacoustic nerve
IX–Glossopharyngeal nerve
X–Vagus nerve
XI–Accessory nerve
XII–Hypoglossal nerve

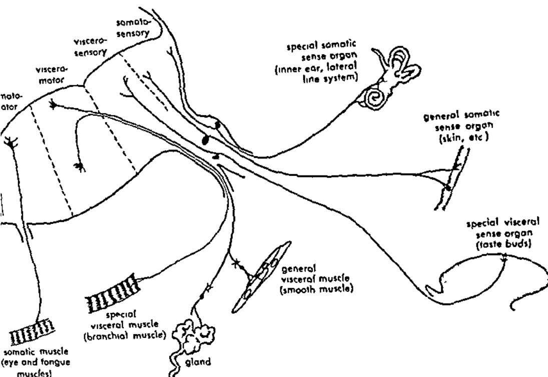


Fig 29 Scheme of the different kinds of nerve fibers in a cranial nerve

Cyclostomes, fishes, and amphibians lack the accessory and hypoglossal nerves

In addition to these nerves a vomeronasal nerve has been described in lower vertebrates and embryonically in man It runs from the so-called Jacobson's organ to the accessory olfactory bulb of the brain and may be regarded as an accessory olfactory nerve Further, a terminal nerve or Pincus's nerve has been described in most vertebrates, although it may lie within the olfactory nerve. It runs from the nasal sac to the ventral surface of the forebrain and contains fibers running to blood vessels and other structures

Functional subdivision The cranial nerves, like the spinal nerves, contain both sensory and motor fibers The former run from sensory end organs to cell bodies situated in the ganglia (the cranial ganglia), continue in the nerve roots, and reach the *brain* stem The motor fibers emanate from cell bodies situated in the brain stem and then run peripherally to *effector organs*

The fibers may be further subdivided into a somatic and a visceral system The *somatic* system in the trunk is represented by the *cross striated* muscles, derived from somites, as *effector organs* and the skin with its derivatives as the sensory system In the trunk the visceral system is made up of the gut and its derivatives, the muscles of which are smooth The approximate correspondences of somatic muscles in the head are represented by the eye muscles and the tongue muscles. The rest of the muscles of the head are derivatives of the gill (branchial) arches and should therefore be regarded as visceral muscles even though they are cross-striated. They are often called special visceral muscles The motor fibers of the cranial nerves, therefore, may be divided into somatomotor fibers (to eye and tongue muscles), general visceromotor fibers (to derivatives of the gut wall), and special visceromotor fibers (to the musculature of the head, derived from the branchial arches). *See* RESPIRATORY SYSTEM.

The sensory fibers are partly of a general somatic nature, innervating the skin and also carrying deep sensibility, and general visceral nature, innervating the gut wall; they run partly from special sense *organs*, typical of the head Thus, special somatic fibers emanate from the *lateral-line* system of aquatic animals and its derivative, *the inner* ear, and special visceral fibers come *from the taste* organ (Fig 29).

The somatomotor fibers in the *head* leave the brain as *ventrally* situated nerve roots; the rest of the fibers make up the dorsally situated roots. Within the brain the *end nuclei* of the different functional systems lie arranged in longitudinal rows or columns

As brain nerves, the nerves of the organs of smell and vision are included also. From a morphologic *point* of view these are, however, not comparable to the rest of the cranial nerves The olfactory nerve is made up of fibers from the primary sensory cells of the olfactory organ and thus does not contain a ganglion The optic nerve is more to be compared to a brain *fascicle*.

Segmentation. The morphology of the cranial nerves is strongly *influenced* by the segmentation present in the *pharynx, that is*, the branchial

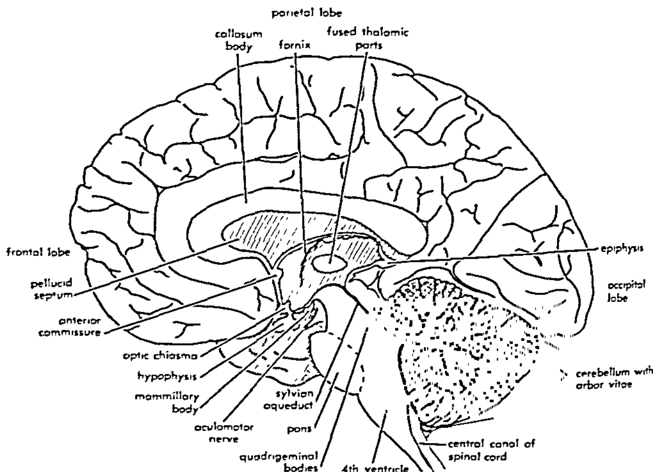


Fig. 28. Drawing of human brain, sectioned medially.

diencephalon with the mammillary bodies, the hypophysis, and the crossing of the optic nerves. On the ventral surface of the two hemispheres, the olfactory bulbs and tracts can be seen (Fig. 27).

A peculiar distribution of white and gray matter of the cerebellum produces the so-called arbor vitae which can be seen in a median section of the brain (Fig. 28). The roof of the mesencephalon appears as the quadrigeminal bodies. Between the two hemispheres lies an enormous commissural bundle, the corpus callosum. Two columnar fiber bundles, the fornices, lie on each side below and behind it connected by a hippocampal commissure. The fornices describe a half-circle from the temporal lobes upward, forward, and downward to end in the mammillary bodies. They form the anterior limit of the foramina of Monro, which connect the ventricles of the hemispheres with that of the diencephalon. Between the corpus callosum and the fornices the brain wall is much extended to form a thin septum, the pellucid septum.

Meninges. The brain is enclosed and protected by three membranes of connective tissue. The outermost one is called the dura mater, and also represents the internal periosteum of the skull in the adult. The innermost one is called the pia mater and is a thin membrane, lying close to the brain surface and extending into the grooves on the brain surface. The arachnoid lies between these two membranes. The space between the arachnoid and the pia is filled by the cerebrospinal fluid, which further protects the brain. This fluid is produced in the brain ventricles from the choroid plexi in the roofs of the third and fourth (myelencephalic) ventricles, fills the ventricular system, leaves it through openings in the myelencephalic tela, and flows into the subarachnoid space, where it is absorbed.

CRANIAL NERVES

The cranial or cerebral nerves are the peripheral nerves of the head, being related to the brain. The number and degree of development of the nerves varies in different species. The functional quality of the different nerves also varies. Twelve pairs of cranial nerves have been distinguished in human anatomy, and these nerves have been numbered rostrally to caudally in the following way:

I–Olfactory nerve, fila olfactoria
II–Optic nerve, fasciculus opticus
III–Oculomotor nerve
IV–Trochlear nerve
V–Trigeminal nerve, in most vertebrates divided into three branches: ophthalmic, maxillary, and mandibular
VI–Abducens nerve
VII–Facial nerve
VIII–Statoacoustic nerve
IX–Glossopharyngeal nerve
X–Vagus nerve
XI–Accessory nerve
XII–Hypoglossal nerve

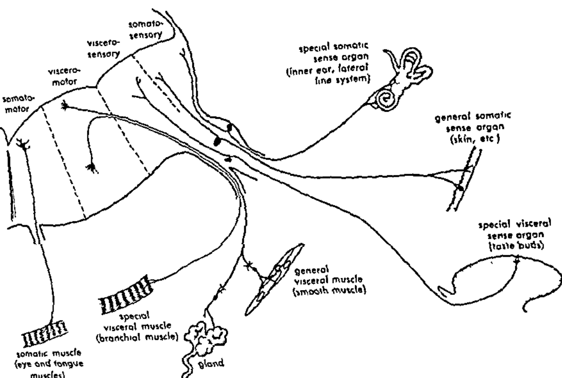


Fig 29 Scheme of the different kinds of nerve fibers in a cranial nerve

Cyclostomes, fishes, and amphibians lack the accessory and hypoglossal nerves

In addition to these nerves a vomeronasal nerve has been described in lower vertebrates and embryonically in man It runs from the so-called Jacobson's organ to the accessory olfactory bulb of the brain and may be regarded as an accessory olfactory nerve Further, a terminal nerve or Pincus's nerve has been described in most vertebrates, although it may lie within the olfactory nerve It runs from the nasal sac to the ventral surface of the forebrain and contains fibers running to blood vessels and other structures

Functional subdivision. The cranial nerves, like the spinal nerves, contain both sensory and motor fibers The former run from sensory end organs to cell bodies situated in the ganglia (the cranial ganglia), continue in the nerve roots, and reach the brain stem The motor fibers emanate from cell bodies situated in the brain stem and then run peripherally to effector organs

The fibers may be further subdivided into a somatic and a visceral system The somatic system in the trunk is represented by the cross-striated muscles, derived from somites, as effector organs and the skin with its derivatives as the sensory system In the trunk the visceral system is made up of the gut and its derivatives, the muscles of which are smooth The approximate correspondences of

... arches and should therefore be regarded as visceral muscles even though they are cross-striated They are often called special visceral muscles. The motor fibers of the cranial

special visceromotor fibers (to the musculature of the head, derived from the branchial arches). *See* RESPIRATORY SYSTEM.

The sensory fibers are partly of a general somatic nature, innervating the skin and also carrying deep sensibility, and general visceral nature, innervating the gut wall, they run partly from special sense organs, typical of the head. Thus, special somatic fibers emanate from the lateral-line system of aquatic animals and its derivative, the inner ear, and special visceral fibers come from the taste organ (Fig 29)

The somatomotor fibers in the head leave the brain as ventrally situated nerve roots, the rest of the fibers make up the dorsally situated roots. Within the brain the end nuclei of the different functional systems lie arranged in longitudinal rows or columns

As brain nerves, the nerves of the organs of smell and vision are included also. From a morphologic point of view these are, however, not comparable to the rest of the cranial nerves The olfactory nerve is made up of fibers from the primary sensory cells of the olfactory organ and thus does not contain a ganglion The optic nerve is more to be compared to a brain fascicle

Segmentation. The morphology of the cranial nerves is strongly influenced by the segmentation present in the pharynx, that is, the branchial

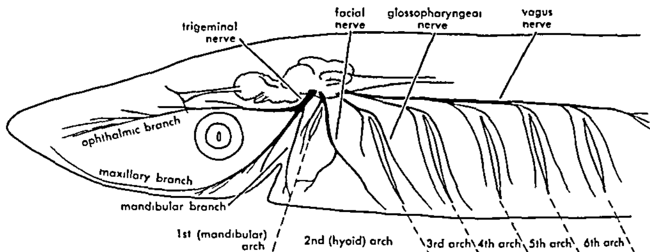


Fig. 30. Branchial nerves of a shark.

arches (branchiomerism) (Fig. 30). In lower vertebrates every nerve root divides and sends one branch (pretrematic branch) in front of a pouch and one behind it (posttrematic branch). In higher forms this pattern may be modified but can still be identified even in man. The ventral roots do not show such a strict branchiomeric arrangement but attempts have been made to combine each ventral root with a dorsal root to form a unit which would fit into a system of branchiomerism and represent a transformed spinal nerve. *See* METAMERISM, EMBRYONIC.

Somatomotor nerves. Three nerves run to the eye muscles: the oculomotor, trochlear, and abducens nerves. The trochlear nerve leaves the brain stem on its dorsal surface after its fibers cross within the brain substance. From a morphologic point of view, however, it is to be regarded as a ventral nerve. The oculomotor and abducens nerves both leave the brain on its ventral surface. The trochlear nerve innervates the superior oblique eye muscle, the abducens nerve the external rectus eye muscle, and the oculomotor nerve the remaining muscles of the eye.

The tongue muscles in man are innervated by the hypoglossal nerve. In fishes the so-called occipital nerves, which leave the spinal cord, pass through the posterior part of the skull, and run to the so-called hypobranchial muscles, correspond to the hypoglossal nerve. In amphibians no occipital nerves are present, but the corresponding nerve trunk is formed from the second and third spinal nerves. The amniote hypoglossal nerve, too, receives nerve fibers from the spinal nerves.

In amniotes another nerve appears which is lacking in lower vertebrates, the accessory nerve. In man it is formed as a specialization of the vagus nerve and also receives a root from the spinal cord. The vagus nerve fibers are special visceromotor ones and enter the vagus ganglion. The spinal cord portion is the true accessory nerve, and its fibers run to the trapezius muscle and the sternocleidomastoideus muscle.

Special visceromotor nerves. These nerves innervate the main part of the head musculature, that derived from the branchial arches. The masticating muscles develop from the first arch (the future lower jaw) and are innervated by the mandibular branch of the trigeminal nerve. The muscles derived from the second branchial arch (hyoid) are innervated by the facial nerve in all vertebrates. In man such muscles are found in the upper part of the neck, and as the stapedius muscle in the middle ear. In amphibians, reptiles, and mammals another muscle develops within the hyoid arch and is innervated by the facial nerve, the platysma muscle. In apes and man it develops into the mimic musculature of the face. The pharyngeal muscles of the third branchial arch are innervated by the glossopharyngeal nerve, and those of the following arches by the vagus nerve.

General somatosensory nerves. The trigeminal nerve contains fibers of this nature in all three branches and innervates the areas situated in front of the lower jaw. The facial nerve in lower vertebrates sends a mandibular branch to the lower jaw which in higher vertebrates and man can be identified as the tympanic cord. The cutaneous innervation in man is represented by a small region in the external ear. The sensory innervation of the third branchial arch is carried through the glossopharyngeal nerve, and that of the following arches through the vagus nerve. In man there is a small remnant of the vagus somatosensory portion as a branch to the external ear.

Special somatosensory nerves. The lateral-line system is a well-developed special sensory organ in the skin of water-living lower vertebrates. It is innervated by the so-called lateral-line nerves which are branches from the trigeminal, facial, glossopharyngeal, and vagus nerves. In amniotes the lateral-line system has disappeared. The inner ear is usually regarded as a special differentiation of this system. Its nerve, the statoacoustic nerve, can therefore be regarded as a remnant of the lateral-line nerves. In amphibians, the tadpoles carry lateral-line nerves which are transformed into a static nerve of the inner ear at metamorphosis.

Special viscerosensory nerves. The nerves from the taste organ in fishes, which extends outside the mouth onto the surface of the head, run as so-called communis fibers in the facial nerve. In amniotes

and man such fibers run through the facial nerve (via the above mentioned tympanic cord), the glossopharyngeal nerve, and the vagus nerve

Visceromotor and sensory nerves. These nerves run to the glands, muscles, and mucous membranes, derived from the gut They represent the cranial parasympathetic system and are made up of fibers in the oculomotor, facial, glossopharyngeal, and vagus nerves

Cranial ganglia. These are made up of the main cranial ganglia, situated in the nerve trunks, and of the parasympathetic ganglia The semilunar ganglion is the main ganglion of the trigeminal nerve, the geniculate ganglion that of the facial nerve. The glossopharyngeal nerve carries two ganglia, the superior and the petrosal ganglia; the vagus nerve also carries two, the jugular and nodose ganglia In some lower vertebrates there is a small dorsal root corresponding to the hypoglossal nerve. Its ganglion is called Froriep's ganglion. It might [illegible]

Bibliography H Bergquist and B Kallén, Notes on the early histogenesis and morphogenesis of the central nervous system in vertebrates, *J Comp. Neurol*, 100(3) 627–659, 1954, O Larsell, *Anatomy of the Nervous System*, 2d ed, 1951, W. von Mollendorf, *Handbuch der Mikroskopischen Anatomie des Menschen*, vol 4, 1957, B M Patten, *Human Embryology*, 2d ed 1953, L S Stone, The development of lateral-line sense organs in amphibians observed in living and vital-stained preparations, *J Comp Neurol.*, 57(3) 507–540, 1933, L S Stone, The origin and development of taste organs in salamanders observed in the living condition, *J Exp. Zool*, 83(3) 481–506 1940, M R. Wright, The lateral line system of sense organs, *Quart. Rev Biol*, 26(3) 264–280 1951, M R Wright, Persistence of taste organs in tongue transplants of *Triturus viridescens*, *J Exp Zool*, 129(2) 357–368 1955

Nervous system (invertebrate)

A nervous system is present in nearly all multicellular animals above the sponges, and in each phylum [illegible] nerve fibers along which signals or nerve impulses are transmitted Nerve cells connect with one another at junctions called synapses across which impulses pass from one cell to another At synapses the surfaces of nerve cells come very close to each other, but there is no fusion of substance Because nerves transmit impulses swiftly, the nervous system is a means of fast regulation Endocrine systems, secreting chemical messengers or hormones into the blood stream, produce slower and more lasting effects. See ENDOCRINE SYSTEM; HORMONE; NERVOUS SYSTEM

Investigations of invertebrate nervous systems have been concerned with basic problems of nervous transmission, and with explaining how animals, or their constituent parts, behave. Knowledge of the functioning of simpler nervous systems often throws light upon events taking place in more complex systems of higher animals, including man. It also gives some insight into the evolution of higher mental processes.

Transmission of a nervous impulse is a physicochemical phenomenon On arriving at the terminal of a fiber, the impulse usually causes the release of a chemical transmitter, which diffuses across the short gap at the synapse and excites the next element in the series, nerve cell, muscle cell, or other structure Nerve fibers that supply muscles, chromatophores, luminous organs, glands, and cilia are known collectively as *effectors* In special instances, nerve cells have been transformed into neurosecretory cells, which discharge hormones from their endings

By controll[illegible] viscera of higher ones, the nervous system takes the form of a diffuse nerve net There is a progressive tendency among higher animals for the nerve cells to be concentrated into a central nervous system. Higher nervous centers, such as ganglia and the brain allow a greater degree of instinctive behavior, more varied responses, and utilization of past experience, that is, memory See BIOLUMINESCENCE, CHROMATOPHORE, MUSCLE

Excitation and transmission. Many kinds of cells can be excited, but this ability is accentuated as the chief function of nerve cells In single-celled animalcules or Protozoa such as *Noctiluca*, action potentials recorded inside the cell, accompany movement of the tentacle or follow electrical stimulation The simple multicellular sponges have no nervous system, yet when they are stimulated, as by a pin prick, excitation is transmitted short distances around the exhalant opening which then closes. Conduction takes place from one contractile musculoepithelial cell to another, and is called neuroid transmission.

The transmission of a nervous impulse along a nerve fiber [illegible] tion [illegible] ducti[illegible] ing [illegible] which take place in an effector when a nerve is stimulated, or directly, by recording the action potential with the aid of an electronic amplifier and cathode-ray oscilloscope

In a resting nerve fiber there is a potential difference across the surface membrane, the inside being negative to the outside By making use of giant nerve fibers found in squid and cuttlefish, it has been possible to introduce an electrode into a fiber and measure the potential inside (Fig. 1) At rest, the potential inside is about 60 millivolts (mv) negative; when the fiber is excited and transmitting an impulse, the potential inside falls to zero and reverses in sign, becoming 50 mv positive; that is, there is a potential change of about 110 mv. The resting potential is caused by differences [illegible]



Fig. 1. Action potential recorded between the inside and outside of the giant axon of a squid The vertical scale shows the potential of the internal electrode in millivolts, the sea water outside being taken as zero potential. The dots below represent time marks at 500 cps. (From A. L. Hodgkin and A F. Huxley, J Physiol., 104(2), 1945)

concentrations of ions across the fiber membrane, the interior of the fiber being rich in potassium ions, and the tissue fluid outside rich in sodium ions

When a nerve fiber is excited, as by an electric shock, a brief change in permeability of the surface membrane occurs, sodium ions rush in and potassium ions pass out These movements of ions produce the action potential—the rising phase of the spike depends on the inward surge of sodium ions; the falling phase, on the outward movement of potassium ions, which restores the resting potential. Current flowing away from the stimulated region excites adjoining regions of the fiber, and so a self-propagating disturbance arises, which proceeds along the fiber to its terminus. By the expenditure of energy, the nerve fiber is able to pump out sodium ions and take in potassium ions, thus restoring its ionic concentrations.

Conduction speed. The speed at which nervous transmission takes place depends on many factors such as the temperature, continuity, and diameter of fibers. The temperature coefficient Q_{10} for nervous transmission is about 1.5; that is, conduction speed increases about 1.5 times for each rise of 10°C. Conduction is retarded by many synapses, each of which imposes a slight delay Large fibers conduct faster than small fibers, the velocity varying approximately as the square root of the diameter. In the nerve nets of coelenterates, consisting of many small nerve cells, impulses are conducted at speeds ranging from 0 04 to 1 meter/sec Small nerves of crabs and squids, which are a few microns in diameter, conduct at speeds of 1–2 m/sec. Giant fibers, up to 1 mm in diameter, conduct at speeds up to 20 m/sec in squid and annelid worms.

Synaptic transmission. Whether judged on morphological or physiological grounds, a wide variety of synapses are encountered among invertebrates At some synapses the areas of contact between two fibers are equal and transmission occurs with equal facility across the junction. Such synapses are said to be nonpolarized. Such synapses occur in through-conduction tracts of sea anemones, between giant axons of serpulid worms, and others. In other synapses, one nerve fiber terminates on another by fine processes, small buttons, or other specialized structures. Transmission is unidirectional, from fine terminals to large fibers, and such junctions are polarized, as exemplified by the sensory terminals on first-order giant fibers in the brain of squid.

From a functional point of view, synapses are further classified as relay or one-to-one, integrative or several-to-one, and multiplying or one-to-several In the relay type each impulse in one fiber (presynaptic) evokes an impulse in the other (postsynaptic), such as the synapse between first- and second-order giant fibers of squid. In the integrative synapse several presynaptic impulses at some optional frequency are necessary to produce postsynaptic discharge, as in the stellate ganglion of the octopus. In multiplying synapses the postsynaptic fiber discharges repetitively to a single impulse in the presynaptic fiber, as occurs in the junction between the central giant fiber and motor fibers in crayfish.

When electrical activity at the junctional region is recorded, the following electrical events are observed Stimulation of the presynaptic fiber produces a propagated presynaptic action potential At the junction a synaptic potential arises. This, an electrotonic potential, is propagated with decrement for very short distances From the synaptic potential arises a propagated all-or-nothing postsynaptic potential or impulse.

The agency or agencies of synaptic transmission among invertebrate animals are still uncertain. Despite the large volume of pharmacological investigation, chemical identity of the transmitter substance has proved elusive Acetylcholine, adrenalin, and 5-hydroxytryptamine have been postulated for various species. There is some evidence, based on sensitivity, that acetylcholine may be involved in neuromuscular transmission in some annelids, mollusks, and echinoderms.

Nerve nets. These structures occur in coelenterates, echinoderms, acorn worms (balanoglossids), and in the gut of annelid worms, crustaceans, and other groups Among the chief physiological characteristics of the nerve net are diffuse spread of excitation, local and equipotential autonomy, spatial attenuation of conduction, and facilitation.

In the sea anemone, for example, the nervous system consists of a diffuse network of nerve cells and processes lying underneath the epithelium. Excitation is conducted in any direction over the anemone from the region stimulated, that is, the nerve tracts are diffuse and nonpolarized. Electrical stimulation of the nervous system evokes one of two kinds of contractions in muscles surrounding the disk. Very slow stimulation produces slow contractions after a long interval, faster stimulation produces stronger contractions The number and frequency of impulses determine whether the muscle will contract slowly or quickly. Usually, several impulses are required to elicit a quick contraction, and subsequent contractions increase progressively in strength.

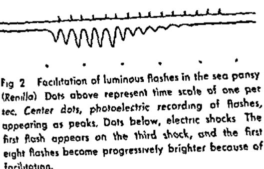


Fig 2 Facilitation of luminous flashes in the sea pansy (*Renilla*) Dots above represent time scale of one per sec. *Center dots, photoelectric recording of flashes,* appearing as peaks. Dots below, electric shocks The first flash appears on the third shock, and the first eight flashes become progressively brighter because of facilitation.

Some excitatory state is carried over from one impulse to the next, and this condition is called facilitation. This process may operate at the junction between nerve and effector cell, or at the synapses between nerve cells Some parts of the nerve net are organized into through-conduction tracts, in which an impulse quickly traverses the whole pathway In others, excitation may spread gradually with continued stimulation, as each impulse paves the way for its successor, allowing it to penetrate further along the nerve trunk (facilitation between nerve cells) In addition to muscular contractions, the nerve net of coelenterates also controls luminous flashes (Fig 2) Despite the apparent structural simplicity of the nerve net, coelenterates display a surprisingly complex gamut of activities

Visceral nervous systems. Intrinsic nerve nets in the alimentary canal of many invertebrates regulate gut movements and glandular secretion Usually they are connected by sympathetic nerves with *brain and nerve* cord, and the activities that *they* control are integrated with the behavior of the whole organism The lugworm *Arenicola*, for example, digs in mud by thrusting out its proboscis at regular intervals This regular periodic activity is regulated by a visceral nerve net in the esophageal wall Excitation originating *in the latter*, radiates out to the proboscis and anterior body wall through nerve tracts, and induces rhythmic muscular contractions

Central nervous systems. This type of system consists of ganglia, or aggregations of nerve cells, linked by conducting tracts or cords Activities involving these centers may be either reflex, that is, called forth by environmental changes, or spontaneous and initiated by the nervous system itself.

A reflex involves sensory stimulation, excitation of efferent nerves, and ...

the state of the animal's body into the nerve centers Here efferent nerves that innervate the effectors are excited, and impulses pass out again to the effector organs; that is, they are reflected back again from the nerve centers to the periphery This is the basis of the reflex arc, and responses invoked in this way by internal or external stimuli are called reflexes.

The reflex concerned with the shortening of an errant polychaete worm can serve as a simple example (Fig 3) Sensory cells in the cirri, or hairlike appendages of the body wall, send nerve fibers into the nerve cord, where they make contact with giant nerve fibers extending through many segments and acting as fast distributors. In each segment the giant fibers have synapses with motor nerve cells, which send fibers to the longitudinal muscles. Touching a cirrus excites the sensory cells, causing them to discharge nervous impulses, which in turn excite the giant fibers, then the motor fibers, and finally the longitudinal muscles.

Th... systems. Parts of the nerv...

innervate extensive muscle masses which they throw into synchronous contraction. Examples are the giant fiber systems controlling the escape reactions of earthworms, lobsters, and squids These extraordinarily large fibers are frequently amalgams of many nerve cells.

The arrangement is fairly simple in annelid worms, in which one or a few giant nerve fibers run throughout the length of the central nervous system from the brain in the head to the tail. In each segment the giant fiber is connected with motor fibers which supply the longitudinal muscles. In one fanworm, *Myxicola*, branches of the giant fiber go directly to the muscles A single nerve impulse rapidly traverses the giant fiber, throwing all the longitudinal muscles into contraction at one time, and causing the worm to shorten suddenly. When a worm is repeatedly stimulated it soon *ceases to contract* Some form of accommodation

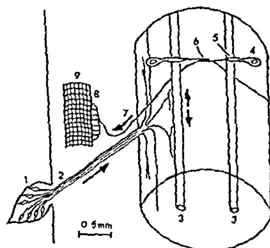


Fig 3 Schematic representation of a reflex arc in an errant polychaete worm, such as Harmothoë 1, cirrus, 2, afferent fibers from sensory cells in the cirrus, 3, pair of giant fibers in the ventral nerve cord, 4, motor nerve cell, 5, synapse of giant fiber and motor fiber, 6, synaptic juxtaposition of motor fibers, 7, motor fiber, 8, contact of motor fiber with muscle, 9, longitudinal muscle. Arrows indicate direction of transmission

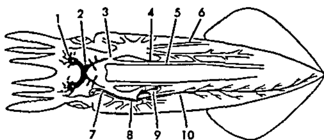


Fig. 4. Giant-fiber system of the squid. Ganglia and nerves enlarged relative to the animal. 1, first-order giant cells; 2, interaxonic bridge; 3, mantle connective nerve; 4, giant-fiber lobe of stellate ganglion; 5, fin nerve, 6, stellar nerve; 7, second-order giant axon; 8, synapse between second- and third-order giant axons; 9, cell bodies of third-order giant axon, 10, third-order axon. (*After J. Z. Young, Cold Spring Harbor Symposia Quant. Biol., vol. 4, 1936*)

takes place in the central nervous system, either between sensory fibers and the giant fiber, or between the latter and motor or efferent fibers, or at both loci. In consequence, synaptic transmission becomes temporarily blocked at these points.

In the squid there is a chain of three giant fibers (Fig. 4), first-, second-, and third-order, extending from the brain to the muscles of the mantle or outer body wall on either side of the body. When the animal is threatened, sensory signals from the eyes or skin excite the first-order giant fibers in the head, from which nerve impulses pass via second- and third-order fibers to the circular muscles of the mantle. The third-order giant fibers are graded in diameter, those going the greatest distances generally being the largest and conducting most quickly Consequently, the muscles contract nearly simultaneously, water is suddenly expelled from the mantle cavity, and the animal shoots away by jet propulsion.

Neurosecretion. In the central nervous systems of many invertebrates there are aggregations of peculiar secretory cells, collectively designated neurosecretory cells. These cells, originally nervous, have become transformed into endocrine cells They are innervated and they respond to nervous stimulation by secreting hormones into the blood stream. Cephalopods illustrate in an interesting manner the transition from nervous to secretory function. In squids, the nerve cell bodies of the third-order giant fibers lie in a special lobe of a peripheral ganglion, the stellate ganglion. In octopuses, however, giant fibers are wanting, and the equivalent nerve cells are represented by neurosecretory cells in a lobe, the epistellar lobe of the stellate ganglion.

Neurosecretory cells are found in brain, nerve cord, and visceral ganglia of nemertines, polychaetes, arthropods, and mollusks Their functioning has been explored most thoroughly in crustaceans and insects. The sinus gland in the eyestalks of crustaceans, for example, is formed of the swollen terminals of fibers coming from neurosecretory cells. In the organs, hormones are stored that, when released, regulate various activities, including molting, respiration, chromatophore responses, and retinal-pigment migration. The triggering factor for hormonal release may be external, as in the case of color changes, when afferent impulses from the eyes enter the brain and eventually reach the neurosecretory centers. Appropriate hormones are released from the neurosecretory terminals and bring about changes in the distribution of pigment in the color cells or chromatophores. Or, in the case of periodic activities such as molting or diurnal changes in respiration, the releasing agency must be internal, originating within the higher nervous centers.

Directive activities. The characteristics of an animal's behavior, spontaneous or provoked by stimulation, are governed by the organization of its higher nervous centers. In these centers incoming signals are sorted out and compared, information is stored, actions are initiated according to the animal's past history and present condition, and rhythmic activities are regulated.

The directive capacities of the nervous system at an elementary level have been revealed in starfish In these animals there is a central nerve ring from which a radial nerve cord arises which passes down each arm. The central ring contains five nerve centers, in each of which there are nerve cells that send fibers down each of the radial nerve cords In the latter, the central fibers contact motor nerves, which supply the muscles of the tube-feet. A starfish moves along by stepping movements of its tube-feet, *one arm leading and the other arms following* During locomotion one of the centers in the nerve ring becomes dominant, coordinates the direction of stepping in all arms, and causes the starfish to move in one particular direction

Many of the activities of animals, although appearing on analysis reflexly modulated, consist of integrated actions involving different organs and taking place in definite temporal sequence. Thus, polychaete worms move in several ways, such as crawling and swimming Crawling is effected by means of peristaltic waves of the body musculature and coordinated movements of the parapodia or appendages The nervous basis for these propagated waves of contraction depends upon several factors. The motor pattern responsible for this particular kind of locomotor activity is propagated and established along the length of the nervous system, timing and sequence of the waves are determined in each segment by chain reflexes evoked by excitation of tactile and tensile receptors or sensory cells in consequence of the movements of contiguous segments

Central versus peripheral control. In some instances control of muscular activity is executed centrally, and the strength of contractions depends on the number of excitatory fibers thrown into activity and the number and frequency of impulses which they carry. More common among inverte-

brates are situations in which muscles receive several kinds of excitatory fibers, each capable of calling forth a different degree of contraction. For example, one of the leg muscles of the cockroach is supplied with a fast and a slow axon. The fast axon produces a fast twitch contraction; the slow axon, a slow, smooth contraction. Crustacean limb muscles receive fast and slow excitatory fibers and, in addition, an inhibitory fiber which inhibits the response of the motor fibers. The characteristics of contraction, then, depend upon the kinds of nerve fibers which are active, and the way in which they are discharging. Resolution of excitatory and inhibitory signals occurs at the periphery.

Functioning of higher centers. The anterior end of bilaterally symmetrical animals, being the region most often exposed to environmental changes, bears the greatest accumulation of sense organs and, simultaneously with these, have appeared the *cephalic or head ganglia*. Primitively, these are relay centers for incoming sensory signals from the head. In the flatworm *Yungia*, for example, the brain is a sensory center of low threshold, impulses from which initiate peristaltic swimming movements. When the brain is removed, as in decapitation, the animal becomes still; stimulation of the cut end renews swimming movements.

Secondarily, *cephalic ganglia* have become association centers, for correlating incoming messages and for storing information, and are the seat of instinctive and plastic behavior.

Even simple animals, such as sea anemones, change their responses to continued stimulation, for instance, repeated feeding. Such altered responses are attributed to adaptation or accommodation in sensory elements, rather than to true learning. The latter involves some lasting change in the [illegible]

[illegible] including conditioned reflexes, occurs in lower animals such as flatworms, annelid worms, and snails. In the earthworm associations are still retained after removal of the cephalic ganglia. The brain becomes dominant in the behavior of higher animals such as crustaceans, insects, and cephalopods.

The brain is very complex in many arthropods and in cephalopods. Much work has been carried out [illegible] together with a visual stimulus, and giving an electric shock, and showing a crab alone, octopuses can be taught to discriminate, and will then attack the crab only when the visual stimulus is absent. The [illegible] trained [illegible] to [illegible] cra[illegible] ple[illegible] has [illegible] me[illegible] lasting several days. Short term memories, lasting a few minutes only, are a function of another region, the optic lobes. The verticalis-lobe system appears to prolong visual memories set up in the optic lobes by presenting these again, from within, and thus causing them to persist for a sufficiently long time to bring about some change of a more permanent nature.

Very little is known about the functioning of the complex insect brain to correlate with the intricate behavioral activities of insects. Peculiar mushroom bodies, the corpora pedunculata, which cap the brain, are thought to be centers in which incoming sensory messages interact in bees and ants. *See* BEHAVIOR, ONTOGENY OF; LEARNING THEORIES.

[J.A.C.N.]

Bibliography: H. Davson, *A Textbook of General Physiology*, 1951; J. A. C. Nicol, *The Biology of Marine Animals*, 1959; G. H. Parker, *The Elementary Nervous System*, 1919; C. L. Prosser (ed.), *Comparative Animal Physiology*, 1950.

Nervous system disorders

Pathology of the nervous system. Knowledge of the functional and structural properties of the normal human nervous tissue is necessary for an understanding of its changes under pathological conditions.

Neuron. The anatomical and functional unit of the nervous system is the neuron. Each neuron consists of a cell body and its processes. Those processes receiving impulses through contact (synapse) with other neurons are called dendrites and the process conveying a nervous impulse to other neurons or end organs is called an axon. The shape of the cell body (nerve cell) varies considerably, and its size ranges from a few to over 100 μ. In the central nervous system (brain and spinal cord) the nerve cells are accumulated in the gray matter whereas their axons are surrounded by fatty substances (myelin), have a white appearance, and form the white matter of the nervous tissue. The nerve cells are located in the outer surface of the brain (cortex) and in the subcortical or basal nuclei; in the spinal cord they are centrally located. In most parts of the cerebral cortex the nerve cells are distributed in seven layers; the number, size, shape, and distribution (cytoarchitecture) vary from cortical field to field depending upon that field's special function. *See* NERVOUS SYSTEM; NUCLEIC ACID.

The nucleus of the nerve cell is usually spherical and contains one or more nucleoli. In the cytoplasm there are granules or clumps that stain blue with basic dyes; this is the Nissl substance or tigroid. The Nissl substance is composed of ribonucleic acid, is sensitive to pathological changes, and therein reveals varying degrees of dissolution (chromatolysis). This phenomenon may be associated with swelling, vacuolization, or ultimately death (liquefaction) of the nerve cell. When the cell bodies show shrinkage, the Nissl substance is condensed and the nucleus becomes pyknotic. Neuro-

fibrils may be demonstrated in the cytoplasm and in the processes of neurons with silver stains.

Glia. In addition to the vessels and connective tissue that are found in all organs of the human body, the central nervous system has another type of supporting tissue known as the glia. The glia is composed of three elements: astrocytes, which are large cells; oligodendroglia cells, which are smaller (both are ectodermal in origin); and microglia cells, which are mesodermal in origin. The exact function of the glial cells under physiological and pathological conditions is still not completely understood. The astrocytes may proliferate (gliosis) and participate in the formation of scars Furthermore, they may undergo reactive and regressive changes such as swelling and shrinkage. They may [illegible] their cyto- [illegible] rable, may become visible using routine stains. Oligodendroglia cells have something to do with the formation and metabolism of myelin. They proliferate in certain toxic diseases and are seen around nerve cells (satellitosis). Microglia cells proliferate very actively in disease and become hypertrophic. They participate in the removal of nuclear and myelin debris (phagocytosis). Focal accumulations of glial cells, predominantly microglia cells, are called glial nodules and are commonly found in viral and rickettsial encephalitides.

Myelin. Pathological changes of myelin may ultimately result in the complete loss of this structure (demyelination). This may occur in infections, intoxications, deficiency diseases (lack of vitamins), anoxia, after destruction of the axon, and for unknown reasons Regardless of the etiology, the degeneration follows a standard scheme which was experimentally studied by interruption of nerve fibers in peripheral nerves (Wallerian degeneration). The Marchi technique stains the altered myelin. Old demyelinated regions do not stain with hematoxylin dyes as does the normal myelin. The axon of a peripheral nerve distal to the injury becomes swollen, fragmented, and undergoes dissolution; the same phenomenon occurs in the myelin structures. Proximal to the injury the myelin becomes degenerated for short distances only and the neurofibrils show proliferation within hours after the injury. When proximal and distal stumps of an injured peripheral nerve are brought together, a regeneration of the myelin and axon is possible and recovery of complete function may follow. On the other hand, no regeneration of neurons, axons, or myelin is possible in the central nervous system

The reaction of vessels, fibroblasts, and collagen in the central nervous system is not different from that in any other organ of the body. These structures respond, along with the glial cells, and participate in all pathological processes of the central nervous system. Their degree of participation in injury, cellular death, repair, and scarring varies from disease to disease.

Meninges. The coverings of the brain and spinal cord are called the meninges; these consist of the pachymeninx or dura and the underlying leptomeninges. The latter are divided into the pia, which is firmly attached to the configurations or gyri of the brain and to the external layer of the spinal cord, and the arachnoid. Between the pia and arachnoid is the subarachnoid space in which the cerebrospinal fluid circulates. Between the dura and arachnoid is the subdural space.

Ventricular system. The cerebrospinal fluid is produced by the choroid plexus and is secreted into the cavities of the brain, the ventricular system The ventricular system consists of four ventricles. The two lateral ventricles communicate with the third ventricle by the foramen of Monro. The communication between the third and fourth ventricles is through the aqueduct of Sylvius. The cerebrospinal fluid leaves the fourth ventricle via the foramina of Luschka and Magendie to reach the subarachnoid space from which it is ultimately absorbed by the Pacchionian granulations The ventricular system of the brain is lined by ependymal cells.

ATROPHY

This term means a decrease in the size or number of individual cells or in the size of a tissue or an organ. Atrophy of one neuron may occur when another functionally and synaptically related neuron is destroyed. This type of atrophy is known as the transsynaptic type of atrophy or degeneration. If many neurons are destroyed, the brain and spinal cord may appear smaller The spinal cord is atrophic in healed cases of poliomyelitis, because this disease destroys the anterior horn cells of the spinal cord. The brain is atrophic or smaller in cases of senile and arteriosclerotic dementia. In the latter the destruction of the neurons is due to a vascular process. In cases of senile dementia, the atrophy of the brain is due to degeneration of the neurons characterized by alterations of the neurofibrils (Alzheimer's neurofibrillary changes). In addition, amorphous argentophilic deposits, the so-called senile plaques, are observed in such cases Neurofibrillary changes and senile plaques are demonstrable with silver carbonate stain In cases of presenile dementia (Alzheimer's disease) the pathological substrate is similar to that of senile dementia; the only difference is the age of the patient (third or fourth decade in presenile dementia). In arteriosclerotic, senile, and presenile dementia the gyri of the brain appear smaller, the sulci larger, and the whole ventricular system is enlarged as a result of the reduction in the amount of nervous tissue. *See* SENILE DEMENTIA

In degenerative processes of the central nervous system the atrophic process does not affect all parts of the brain equally, but has a certain predilection for functional regions or functional systems of the cortex, basal nuclei, or spinal cord This atrophy is symmetrical and affects both sides of the central nervous system Most of these diseases are hereditary. For unknown reasons the nerve cells undergo a slow degenerative process covering a

span of many years and ultimately leading to the complete destruction of the neurons. In Pick's disease atrophy predominantly occurs in the frontal lobes of the brain, often associated with atrophy of the temporal or parietal lobes In Huntington's chorea the process affects predominantly the caudate nucleus In Friedreich's ataxia and spinocerebellar atrophies the nervous tissue of the spinal cord and of the cerebellum undergoes progressive degeneration In amyotrophic lateral sclerosis the motor cortex of the brain, the anterior horns of the spinal cord, and often the bulbar nuclei are affected In spinal atrophies the destruction of the *anterior* horn cells causes secondary degeneration of the innervated muscles of the extremities. *See* HUNTINGTON'S CHOREA

MALFORMATIONS

These diseases are due to developmental defects originating in the prenatal and postnatal life.

Hydrocephalus. The most common malformation is the congenital hydrocephalus which results from lack of communication of the cerebrospinal fluid between the ventricular system and the subarachnoid space over the convexity of the brain The ventricular system is maximally enlarged because of accumulation of cerebrospinal fluid This causes pressure upon and atrophy of the brain Obstruction of the circulation of the cerebrospinal fluid may occur at the level of the aqueduct of Sylvius or in the foramina of Luschka and Magendie (noncommunicating hydrocephalus) Often the obstruction of the circulation of the cerebrospinal fluid is around the subarachnoid spaces of the cerebellum, this is called communicating hydrocephalus because the cerebrospinal fluid has an exit from the ventricular system to the subarachnoid space of the cerebellum and spinal cord, but cannot gain access to the subarachnoid space over the convexity of the brain where it may be absorbed

Spinal cord malformations. Often congenital hydrocephalus is associated with a malformation of the spinal cord The bony structures of the spine may not close at birth (spina bifida); the meninges may herniate out of the bony defect (meningocele) or the meninges plus the spinal cord may protrude out of the defect (meningomyelocele). Similar bony defects with herniation of the meninges and the brain tissue may occur in the occipital region (meningoencephalocele). In the Arnold-Chiari malformation the cerebellum and medulla are present in the enlarged upper cervical cord and there is usually an associated spina bifida and hydrocephalus, the bony structures in the posterior fossa are shallow (platybasia)

Other malformations. Other types of malformation are characterized by a very small brain (microcephalus) or a very large brain (megalocephalus) or the configuration of the brain may appear too small (microgyria) or too broad (pachygyria) Regions of the central nervous system may fail to develop (agenesis), for example, the cerebellum may be absent. In some rare instances only rudimentary brain tissue may be present (anencephaly), associated with agenesis of most of the bones of the calvaria. In porencephaly cavities extend from the surface of the cerebral hemisphere to the ventricular system In another malformation, syringomyelia, there is characteristic tubular cavitation of the spinal cord; this process may also affect the medulla (syringobulbia). Tuberous sclerosis (Bourneville's disease) is characterized by focal proliferation and accumulations of abnormal cells which cannot be defined either as nerve cells or as glial elements. Tumors of the heart, kidneys, and other visceral organs have been described associated with this disease. In addition, cutaneous abnormalities and bony defects are present.

Mongolism is one of the most common malformations of the central nervous system. In this disease the brain may be grossly and microscopically normal. There are approximately 300,000 mongolian idiots in the United States and the individuals affected have mongoloid features The cause of this malformation is not known; however, chromosomal aberrations have recently been reported in mongolian idiots In addition to the neurological disorder, some mongolian idiots have abnormalities of the heart In some instances familial occurrence of this disease has been [illegible]

idiocy [illegible]

occurs [illegible]

dividuals affected are blind (amaurosis) and have progressive muscular weakness and subnormal mental development The characteristic histologic feature of this condition is an abnormal accumulation of a lipid, lecithin, in the cytoplasm of the nerve cells Degeneration of the neurons bearing the lipid is followed by scarring (gliosis) Spastic diplegia (Little's disease) which is characterized by spasticity of the lower extremities is not due to the same cause as amaurosis. This disease offers a variety of pathological findings, such as degenerative and atrophic changes, as well as malformation of the brain *See* MONGOLIAN IDIOCY.

The cause of malformations of the central nervous system is obscure. In some rare instances, infections, such as German measles of the pregnant mother, are the cause of malformations. Experimentally, cerebral malformations in newborn animals can be produced by x-rays or by anoxia of the pregnant mother. *See* TERATOGENESIS.

INFLAMMATORY DISEASES

Meningitis, encephalitis, and myelitis are terms which imply inflammatory diseases of the brain coverings (meningitis), the brain (encephalitis), or of the spinal cord (myelitis). The causative agents vary; they may be parasites, bacteria, fungi, rickettsiae, or viruses.

Meningitis. The term leptomeningitis or meningitis refers to an inflammation of the pia-arachnoid in which the exudate accumulates in the subarachnoid space. The exudate may be composed

predominantly of polymorphonuclear leukocytes and fibrin (purulent meningitis) or predominantly of lymphocytes (nonpurulent meningitis). Purulent meningitides are due to meningococci, pneumococci, streptococci, staphylococci, gonococci, influenza B bacillus, actinomycetic infections, and also often by mixed infections. Often the process extends into the ventricular system or into the underlying brain tissue (meningoencephalitis). So-called epidemic meningitis is a purulent meningitis caused by meningococci; very severe cases are also characterized by the presence of petechiae (hemorrhages of the skin). Examples of nonpurulent meningitides are the tuberculous, syphilitic, torula, and coccidioidal meningitides and lymphocytic choriomeningitis. All these meningitides are caused by bacteria or fungi, with the exception of lymphocytic choriomeningitis which is caused by a virus. *See* HEMORRHAGE.

Encephalitis. This inflammatory condition may be caused by bacteria, parasites, fungi, rickettsiae, or viruses. An encephalitis may be associated with inflammation of the meninges (meningoencephalitis) or of the spinal cord (encephalomyelitis) or of both (meningoencephalomyelitis). The lesions of encephalitis may occur predominantly either in the gray matter (polioencephalitis) or in the white matter (leukoencephalitis).

Bacteria. An example of a bacterial infection of the brain is the so-called metastatic or embolic encephalitis, characterized by the presence of small miliary abscesses in the brain. This disease represents a secondary infection of the brain through infectious emboli following primary infectious diseases of the lung, such as pneumonia-empyema, or of the heart, as in subacute bacterial endocarditis. The bacteria causing the abscesses can be demonstrated in the brain, and the exudate is predominantly composed of polymorphonuclear leukocytes (purulent encephalitis). The meninges reveal similar inflammatory changes to the brain (purulent meningoencephalitis). A similar infection of the brain may occur after trauma and in all instances in which the nervous tissue is reached with the blood stream by infectious material. Large solitary abscesses may occur in the central nervous system after trauma, otitis media, or other conditions. There are encephalitides in which granulomas are prominent histological findings (granulomatous encephalitis). Thus, in tuberculosis and syphilis, tuberculomas and gummas, respectively, may be found in the central nervous system. In both diseases the histological characteristics of the granulomas of the brain are similar to the tuberculomas or gummas occurring elsewhere.

Parasites. Examples of parasitic encephalitides are trypanosomiasis or African sleeping sickness, toxoplasmosis, and trichinosis. Infections of the nervous system caused by cysticercus and by malaria organisms may occur. *See* CYSTICERCOSIS; MALARIA; SLEEPING SICKNESS, AFRICAN; TOXOPLASMOSIS; TRICHINOSIS.

Rickettsia. In rickettsial encephalitides such as epidemic typhus, scrub typhus, and Rocky Mountain spotted fever the most characteristic histological lesions are the so-called glial nodules or typhus nodules. These are composed of lymphocytes, perithelial cells, plasma cells, and macrophages, as well as microglia cells. Occasionally polymorphonuclear leukocytes are present. In Rocky Mountain spotted fever these glial nodules occur predominantly in the white matter. In rickettsial encephalitides changes in the capillaries and small arteries are present. *See* SPOTTED FEVER, ROCKY MOUNTAIN; TYPHUS FEVER, EPIDEMIC (LOUSE-BORNE); TYPHUS, SCRUB.

Viral encephalomyelitides. In most instances a virus has been isolated in these diseases. In epidemic encephalitis (encephalitis A, lethargic encephalitis, or von Economo encephalitis) a viral agent was never demonstrated. However, the involvement of the nervous tissue and the histologic changes of this disease are so characteristic for a viral encephalitis that a viral etiology is suspected. The same holds true for the inclusion encephalitis. In this condition characteristic intranuclear inclusions, similar to those seen in herpes encephalitis, are observed in glial and nerve cells.

There are histologic changes common to all viral encephalitides. One of the earliest and most prominent features is the destruction of the nerve cells. Later the affected areas are infiltrated with polymorphonuclear leukocytes. At the height of the disease, proliferating microglia cells and lymphocytes replace the vanishing polymorphonuclear leukocytes. Thick lymphocytic infiltrates are seen around vessels. Microglia cells and lymphocytes form circumscribed lesions, the glial nodules. The dead neurons are either dissolved or phagocytized by polymorphonuclear leukocytes and later by microglia cells (neuronophagia). Some characteristic features may be found in a few viral encephalitides. Thus, in rabies, cytoplasmic inclusion bodies (Negri bodies) enable the histologist to diagnose this disease more specifically.

Intranuclear inclusions are encountered in viral infections such as inclusion encephalitis and herpes simplex encephalitis. Although the nature of the lesions is always the same in viral encephalitides, their distribution in brain and spinal cord is different from one patient to another. This phenomenon is helpful in diagnosing and classifying these infections of the nervous system. According to the distribution of the lesions A. Wolf classifies the viral encephalitides into the following three groups. In the first group the involvement occurs pre- [illegible]

and the basal nuclei; practically no part of the central nervous system is spared in these encephalitides. In this group are included the Japanese B encephalitis (the second type of encephalitis occurring in Japan; the first encephalitis described

in Japan was the epidemic encephalitis or encephalitis A), the Eastern and Western equine encephalitides, the St. Louis encephalitis, the Venezuelan encephalitis, and the Russian spring-summer encephalitis. In the third group of encephalitides, the process is predominantly located in the cerebral cortex. Examples of these diseases are the herpes encephalitis and the inclusion encephalitis; the latter may occur as an acute, subacute, or chronic type, often lasting for months In the first and second groups the spinal cord is usually affected and the lesions are located exactly as in poliomyelitis, in the anterior horn cells Therefore, it is impossible to distinguish one encephalitis from another on the basis of an examination of the spinal cord alone [illegible]

[illegible] CEPHALITIDES.

Poliomyelitis. The most common and most important viral infection of the central nervous system is poliomyelitis (infantile paralysis or Heine-Medin disease) In this condition the nerve cells of the anterior horns in the spinal cord suffer *the brunt* of the attack The lumbar region of the spinal cord *is more* often affected The disease may severely attack the medulla (bulbar polio) Important nervous control centers such as the center controlling respiration are located in the medulla In all cases of poliomyelitis there is some involvement of the brain This is usually [illegible] cortex; [illegible] *vermis* [illegible] are als[illegible] contact [illegible] *polio* virus develop paralysis. Some people have as the only manifestation of this contact antibodies in their blood against the virus, whereas others for unknown reasons develop poliomyelitis. See POLIOMYELITIS

For a long time it was thought that the viruses causing the viral encephalitides were able to [illegible]

The postinfectious or perivenous encephalitis may develop as a complication during and shortly after measles, chickenpox, or mumps, and after vaccination with cowpox or rabies vaccine This encephalitis is characterized by perivenous foci of demyelination and infiltrates which are composed predominantly of microglia cells.

Syphilitic infections. These infections may be manifested in different forms in the nervous system It has already been mentioned that syphilis may cause a nonpurulent leptomeningitis or granulomatous encephalitis Syphilis may affect the vessels of the brain, causing inflammatory changes and thickening of their walls (syphilitic arteritis or endarteritis). Further manifestations of tertiary syphilis in nervous tissue are general paresis or paralysis and tabes dorsalis. *General* paresis is an encephalitis characterized by perivascular infiltrates composed of lymphocytes and plasma cells and by proliferation of the microglia cells (rod cells). The lesions are present in the cortex of the brain where the nerve cells *undergo* degenerative changes leading to their destruction In tabes dorsalis a degeneration of the posterior columns of the spinal cord is seen, associated with demyelination and subsequent gliosis. See PARESIS, GENERAL.

DEMYELINATING DISEASES

In this group are the rare cases of diffuse leukodystrophy or Schilder's disease and the important disease disseminated sclerosis or *multiple sclerosis.*

Multiple sclerosis. There is hardly an agent that has not been held responsible for multiple sclerosis which is, after the cerebral vascular accidents, the most common neurological disease At present nothing is known about the etiology of *multiple sclerosis, which has a chronic* course with periods of remissions and exacerbations. Multiple sclerosis is *characterized* by lesions of varying size *in* which the myelin is destroyed; in such areas the axons are *more resistant to* the disease and may be preserved In older foci the destroyed myelin is replaced by glial scars which appear brown and firm Many inflammatory cells (lymphocytes and plasma cells) *may be seen in* recent lesions. Therefore some investigators believe that multiple sclerosis belongs *to the* group of encephalitides and *may* be called disseminated encephalomyelitis with demyelination There is no region of the central nervous system which cannot be affected *in multiple sclerosis* In some rare instances multiple sclerosis may run an acute fulminant course *leading* to the death of the patient (acute disseminated sclerosis) within a few weeks' time

Extrapyramidal motor system. Two prominent diseases of the extrapyramidal motor system *are* Wilson's disease (hepatolenticular degeneration) and Parkinson's disease (paralysis agitans) In the diseases *of the* extrapyramidal system the basal *nuclei are* predominantly affected In Wilson's disease a necrosis of the putamen and caudate *nucleus (lenticular* nucleus) occurs, associated with cirrhosis of the liver In addition, a greenish *pigmentation* is present along the corneal margins of the eye (Kayser-Fleischer *ring*) Hepatolenticular degeneration *is due* to a disturbance of copper metabolism. Paralysis *agitans is* a progressive disease characterized by coarse tremors and muscular rigidity Degenerative changes of the neurons are present in the basal nuclei and in the substantia *nigra, which* is markedly *degenerated* in this condition.

Diseases that affect the central nervous system primarily have been discussed in this *article.* In addition, the nervous tissue is *affected* by many diseases occurring *in the* other organs *of the* body. In many *infections* and intoxications the *central* nervous system reacts with degeneration of the

neurons and changes of the supporting tissues, both glial and mesenchymal. Many of the diseases of the peripheral nervous system (polyneuritides or polyneuropathies) are secondary to generalized diseases of the human body. In diseases of the liver such as cirrhosis, changes of the glial cells are described. In erythroblastosis fetalis which is associated with icterus, an icteric discoloration of the basal nuclei (kernicterus) is observed in newborns suffering from this disease. In pernicious anemia, a degeneration of the white matter of the spinal cord may occur (subacute combined degeneration). These are only a few examples indicating that a close relationship exists between the central nervous system and the other organs of the body. *See* LIVER DISORDERS. [E.E.M.]

Bibliography: J. G. Greenfield, W. Blackwood, W. H. McMenemy, A. Meyer, and R. M. Norman, *Neuropathology*, 1958; J. G. Kidd (ed.), *The Pathogenesis and Pathology of Viral Diseases*, 1950; E. L. Potter, *Pathology of the Fetus and the Newborn*, 1952.

Network theory, electrical

Network theory includes both analysis and synthesis. In analysis, the network is given, and the problem is to determine the response for specified excitation functions or to determine the excitation function for a specified response. In synthesis, the response and excitation functions are specified, and the problem is to find a network which will satisfy the prescribed conditions.

In both analysis and synthesis, it is important to have a good understanding of circuit laws and network theorems. This article describes the more common and, perhaps, more important laws and theorems.

Ohm's law. In dc circuits, Ohm's law states that the applied voltage V equals the product of the current I and the resistance R

$$V = RI \tag{1}$$

In ac circuits, it is common practice to consider Ohm's law as

$$\mathbf{V} = \mathbf{ZI} \tag{2}$$

where $\mathbf{Z}$ is the impedance of the circuit and all quantities are phasors *See* ALTERNATING-CURRENT CIRCUIT THEORY; OHM'S LAW.

Kirchhoff's laws. Kirchhoff's laws dealing with electric circuits may be stated as follows:

Voltage law The phasor sum (or algebraic sum for instantaneous values or dc values) of voltages around any closed loop of a network is zero.

Current law. The phasor sum (or algebraic sum for instantaneous values or dc values) of the currents at any junction point of a network is zero. *See* KIRCHHOFF'S LAWS OF ELECTRIC CIRCUITS.

Methods of solution. A network may be solved when it is possible to set up a number of independent equations equal to the number of unknown quantities. Therefore, to solve circuit problems it is necessary to be able to identify the specific physical principles involved in the problem and to use these principles to write equations expressing the relations among the unknowns. These expressions may be differential equations or steady-state equations The form used will depend on the particular problem and the results desired.

For simplicity, this article primarily discusses the setting up of the steady-state equations. The methods of setting up the differential equations are similar to those of setting up the steady-state equations. Examples are given to illustrate this. For clearness, all generators in the same network are considered to have the same frequency.

Ohm's law and Kirchhoff's laws furnish the basic information for writing the independent equations. Two common procedures for writing the independent equations are the loop method and the node method. Essentially, the loop method uses loop voltage equations while the node method uses node current equations. The method used is usually the one which has the fewer simultaneous equations to solve.

Network topology. The mathematical science of elements and connections is known as topology. This branch of mathematics gives useful information on how to obtain the necessary number of independent equations

When both the branch currents and the branch voltages are unknown, a network of b branches will require $2b$ equations. The b branch currents may be determined from a combination of loop equations and node equations. The b branch voltages may then be found by applying $\mathbf{V} = \mathbf{ZI}$ to each branch.

For simplicity, assume that only currents are required This requires b equations. Topology determines the proper number of loop equations and the proper number of node equations for finding the currents In Fig 1 the branch currents (the current flowing through a branch, such as *be* or *eab*) are I_1, I_2, I_3, I_4, I_5 and the major nodes (a junction of more than two branches) are b, c, e.

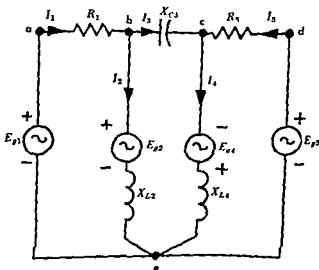


Fig 1.

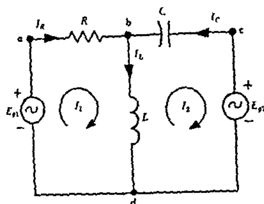


Fig 2

If n_m is the number of major nodes, $n = (n_m - 1)$ is the number of independent major nodes, using one major node as reference, and l is the number of independent loops, then

$b = l + n$ = number of independent equations (3)

$l = (b - n)$ = number of independent loop equations (4)

$n = (n_m - 1)$ = number of independent node equations (5)

In a network diagram which is properly marked, it is easy to count the number of branches to determine b and to count the number of major nodes n_m to find $n = (n_m - 1)$ The number of independent loops is $l = (b - n)$

Loop method. The loop method consists of writing only the l loop voltage equations with loop currents. The solution for the loop currents will give the branch currents by comparison of current arrows in the circuit diagram

If, in Fig 2, the solution for the three branch currents I_R, I_L and I_C is required the loop method may be used to find the loop currents I_1 and I_2, and the branch currents may be found from the relations

$$\mathbf{I}_R = \mathbf{I}_1 \qquad \mathbf{I}_C = -\mathbf{I}_2 \qquad \mathbf{I}_L = \mathbf{I}_1 - \mathbf{I}_2 \qquad (6)$$

The loop method may be *outlined* in the following steps

Step 1 Determine the *number* of independent loop equations

$$l = b - n = 3 - 1 = 2 \qquad (7)$$

Step 2. Arbitrarily choose the loop currents I_1 and I_2 and write the independent loop equations

$$\mathbf{E}_{g1} = (R + jX_L)\mathbf{I}_1 - jX_L\mathbf{I}_2 \qquad (8)$$

$$-\mathbf{E}_{g2} = -jX_L\mathbf{I}_1 + j(X_L - X_C)\mathbf{I}_2 \qquad (9)$$

Step 3. Solve the loop equations simultaneously by the method of determinants. Thus

$$\mathbf{I}_1 = \frac{j(X_L - X_C)\mathbf{E}_{g1} - jX_L\mathbf{E}_{g2}}{X_LX_C + jR(X_L - X_C)} \qquad (10)$$

$$\mathbf{I}_2 = \frac{-(R + jX_L)\mathbf{E}_{g2} + jX_L\mathbf{E}_{g1}}{X_LX_C + jR(X_L - X_C)} \qquad (11)$$

Step 4 Find the branch currents by the relations given in Eq (6).

Judgment in choosing the loop currents can reduce the algebraic work. For example, if only I_L in Fig. 2 is required, then I_1 could be chosen as shown and I_2 chosen around the outer loop *cdabc*. Then $\mathbf{I}_L = \mathbf{I}_1$ and it would not be necessary to find $\mathbf{I}_2$.

To find the instantaneous loop currents i_1 and i_2 in Fig. 2, the procedure is the same as in the steady-state cases, except that differential equations are required

$$e_{g1} = Ri_1 + L\frac{d(i_1 - i_2)}{dt} = Ri_1 + L\frac{di_1}{dt} - L\frac{di_2}{dt}$$

$$-e_{g2} = L\frac{d(i_2 - i_1)}{dt} + \frac{1}{C}\int i_2\,dt$$

$$= -L\frac{di_1}{dt} + L\frac{di_2}{dt} + \frac{1}{C}\int i_2\,dt$$

Figure 3 shows a network with two known ideal *current* generators. The number of independent *loop equations* is

$$l = b - n = 5 - 2 = 3$$

Since $\mathbf{I}_1 = \mathbf{I}_a$ and $\mathbf{I}_3 = \mathbf{I}_b$ are known, then only one loop equation will be needed to find $\mathbf{I}_2$. To find the instantaneous value

$$0 = L\frac{d(i_2 - i_1)}{dt} + Ri_2 + \frac{1}{C}\int (i_2 - i_3)\,dt$$

$$= L\frac{d(i_2 - i_a)}{dt} + Ri_2 + \frac{1}{C}\int (i_2 - i_b)\,dt$$

The steady-state current may be found as follows:

$$0 = j\omega L(\mathbf{I}_2 - \mathbf{I}_1) + R\mathbf{I}_2 - j\frac{1}{\omega C}(\mathbf{I}_2 - \mathbf{I}_3)$$

$$= j\omega L(\mathbf{I}_2 - \mathbf{I}_a) + R\mathbf{I}_2 - j\frac{1}{\omega C}(\mathbf{I}_2 - \mathbf{I}_b)$$

Sometimes it is convenient to write the loop equations in the so-called standard form with loop currents This will be useful in the analysis and solution of network problems.

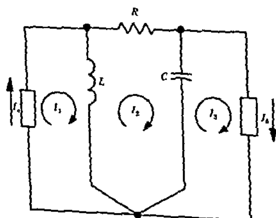


Fig 3.

neurons and changes of the supporting tissues, both glial and mesenchymal. Many of the diseases of the peripheral nervous system (polyneuritides or polyneuropathies) are secondary to generalized diseases of the human body. In diseases of the liver such as cirrhosis, changes of the glial cells are described. In erythroblastosis fetalis which is associated with icterus, an icteric discoloration of the basal nuclei (kernicterus) is observed in newborns suffering from this disease. In pernicious anemia, a degeneration of the white matter of the spinal cord may occur (subacute combined degeneration). These are only a few examples indicating that a close relationship exists between the central nervous system and the other organs of the body. *See* LIVER DISORDERS. [E.E.M.]

Bibliography: J. G. Greenfield, W. Blackwood, W. H. McMenemy, A. Meyer, and R. M. Norman, *Neuropathology*, 1958; J. G. Kidd (ed.), *The Pathogenesis and Pathology of Viral Diseases*, 1950; E. L. Potter, *Pathology of the Fetus and the Newborn*, 1952.

Network theory, electrical

Network theory includes both analysis and synthesis In analysis, the network is given, and the problem is to determine the response for specified excitation functions or to determine the excitation function for a specified response. In synthesis, the response and excitation functions are specified, and the problem is to find a network which will satisfy the prescribed conditions.

In both analysis and synthesis, it is important to have a good understanding of circuit laws and network theorems. This article describes the more common and, perhaps, more important laws and theorems

Ohm's law. In dc circuits, Ohm's law states that the applied voltage V equals the product of the current I and the resistance R

$$V = RI \tag{1}$$

In ac circuits, it is common practice to consider Ohm's law as

$$\mathbf{V} = \mathbf{ZI} \tag{2}$$

where **Z** is the impedance of the circuit and all quantities are phasors *See* ALTERNATING-CURRENT CIRCUIT THEORY; OHM'S LAW.

Kirchhoff's laws. Kirchhoff's laws dealing with electric circuits may be stated as follows:

Voltage law. The phasor sum (or algebraic sum for instantaneous values or dc values) of voltages around any closed loop of a network is zero

Current law The phasor sum (or algebraic sum for instantaneous values or dc values) of the currents at any junction point of a network is zero. *See* KIRCHHOFF'S LAWS OF ELECTRIC CIRCUITS.

Methods of solution. A network may be solved when it is possible to set up a number of independent equations equal to the number of unknown quantities. Therefore, to solve circuit problems it is necessary to be able to identify the specific physical principles involved in the problem and to use these principles to write equations expressing the relations among the unknowns. These expressions may be differential equations or steady-state equations The form used will depend on the particular problem and the results desired.

For simplicity, this article primarily discusses the setting up of the steady-state equations The methods of setting up the differential equations are similar to those of setting up the steady-state equations. Examples are given to illustrate this For clearness, all generators in the same network are considered to have the same frequency.

Ohm's law and Kirchhoff's laws furnish the basic information for writing the independent equations. Two common procedures for writing the independent equations are the loop method and the node method Essentially, the loop method uses loop voltage equations while the node method uses node current equations. The method used is usually the one which has the fewer simultaneous equations to solve.

Network topology. The mathematical science of elements and connections is known as topology This branch of mathematics gives useful information on how to obtain the necessary number of independent equations.

When both the branch currents and the branch voltages are unknown, a network of b branches will require $2b$ equations The b branch currents may be determined from a combination of loop equations and node equations The b branch voltages may then be found by applying $\mathbf{V} = \mathbf{ZI}$ to each branch.

For simplicity, assume that only currents are required. This requires b equations Topology determines the proper number of loop equations and the proper number of node equations for finding the currents In Fig 1 the branch currents (the current flowing through a branch, such as be or eab) are I_1, I_2, I_3, I_4, I_5 and the major nodes (a junction of more than two branches) are b, c, e

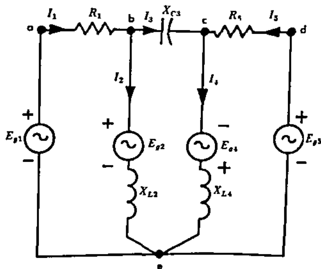


Fig. 1.

node equations are needed to determine the currents.

Step 2. Since the major node e is used as the reference, the node equations at the major nodes b and c are

$$\frac{\mathbf{V}_a - \mathbf{V}_b}{R_1} = \frac{\mathbf{V}_b}{j\omega L_1} + \frac{\mathbf{V}_b - \mathbf{V}_c}{-j\dfrac{1}{\omega C}} \qquad (22)$$

$$\frac{\mathbf{V}_b - \mathbf{V}_c}{-j\dfrac{1}{\omega C}} + \frac{\mathbf{V}_d - \mathbf{V}_c}{R_2} = \frac{\mathbf{V}_c}{j\omega L_2} \qquad (23)$$

Step 3. Since, by inspection of Fig 4, $\mathbf{V}_a = \mathbf{E}_{g1} = 100/\underline{0°}$ and $\mathbf{V}_d = \mathbf{E}_{g2} = 50/\underline{90°}$, then (22) and (23) become

$$\frac{100/\underline{0°} - \mathbf{V}_b}{5/\underline{0°}} = \frac{\mathbf{V}_b}{10/\underline{90°}} + \frac{\mathbf{V}_b - \mathbf{V}_c}{10/\underline{-90°}} \qquad (24)$$

$$\frac{\mathbf{V}_b - \mathbf{V}_c}{10/\underline{-90°}} + \frac{50/\underline{90°} - \mathbf{V}_c}{10/\underline{0°}} = \frac{\mathbf{V}_c}{10/\underline{90°}} \qquad (25)$$

The expressions of (24) and (25) are two independent node equations with two node voltages $\mathbf{V}_b$ and $\mathbf{V}_c$ as unknowns. For description of the complex notations used here, *see* ALTERNATING-CURRENT CIRCUIT THEORY

Step 4. Solving (24) and (25) simultaneously gives

$$\mathbf{V}_b = 69.5/\underline{-19.4°} \qquad \mathbf{V}_c = 83.33/\underline{123.8°}$$

Step 5. Substitution of values in (17) gives

$$\mathbf{I}_{R1} = \frac{100/\underline{0°} - 69.5/\underline{-19.4°}}{5/\underline{0°}} = 6.92 + j4.61$$

$$\mathbf{I}_{L1} = \frac{69.5/\underline{-19.4°}}{10/\underline{90°}} = -2.31 - j6.53$$

$$\mathbf{I}_C = \frac{69.5/\underline{-19.4°} - 83.33/\underline{123.8°}}{10/\underline{-90°}} = 9.23 + j11.14$$

$$\mathbf{I}_{L2} = \frac{83.33/\underline{123.8°}}{5/\underline{90°}} = 13.84 + j9.22$$

$$\mathbf{I}_{R2} = \frac{50/\underline{90°} - 83.33/\underline{123.8°}}{10/\underline{0°}} = 4.61 - j1.92$$

Judgment in selecting the reference node may reduce algebraic work. For example, if only $\mathbf{I}_C$ in Fig. 4 is required, then using node c (a node at either end of the branch) as reference requires that only $\mathbf{V}_b$ need be found for determining

$$\mathbf{I}_C = \frac{\mathbf{V}_b}{-j(1/\omega C)}$$

Y and Δ sections. Figure 5*a* shows a Y section in power circuits. In communication networks, this is more commonly called a T section and is arranged as shown in Fig. 5*b*. Figure 6*a* gives a Δ section in power circuits; while Fig. 6*b* shows the

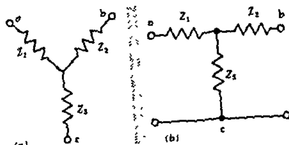


Fig. 5. (a) Y section. (b) T section.

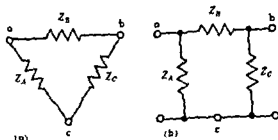


Fig. 6. (a) Δ section. (b) π section

Δ section as a π section in communication networks

The Y or T section and the Δ or π section are sometimes classified as three terminal networks

In the analysis and solution of a network, it is sometimes desirable to transform a Y section into an equivalent Δ section or vice versa. For methods of transformation *see* Y-DELTA TRANSFORMATIONS.

Network theorems. In the solution of specific problems, much time can often be saved by making use of special methods or relations, known as network theorems, instead of starting with Ohm's law and Kirchhoff's voltage and current laws. In analysis, network theorems often give a better insight into the behavior of networks.

There are many network theorems. To use some of these theorems, it is necessary only to recognize certain fundamental similarities between new complicated structures and other established simpler networks. Some of the more common and useful theorems are introduced here.

Superposition theorem. In any linear, bilateral network containing generators, the current flowing in any branch is the sum of the currents which would result from each generator acting independently, while all other generators are replaced at the time by their internal impedances.

The principle of superposition, one of the most important theorems in network analysis, is also useful in proving other theorems. Other theorems are more useful for numerical calculations. *See* SUPERPOSITION THEOREM (ELECTRIC NETWORKS)

Thévenin's theorem. The current in any impedance $\mathbf{Z}_L$, between terminals a-b of a linear, bilateral network containing generators of the same frequency, is equal to the current in the same $\mathbf{Z}_L$ when it is connected to a voltage generator whose generated voltage is the voltage at terminals a-b

Two-loop network. For the two-loop network in Fig. 2, the loop equations (8) and (9) may be written in the standard form

$$\mathbf{E}_1 = \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 \quad (12)$$
$$\mathbf{E}_2 = \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 \quad (13)$$

where $\mathbf{E}_1 = \mathbf{E}_{g1}$ = total emf in loop 1

$\mathbf{E}_2 = -\mathbf{E}_{g2}$ = total emf in loop 2

$\mathbf{I}_1$ = loop current in loop 1

$\mathbf{I}_2$ = loop current in loop 2

$\mathbf{Z}_{11} = R + jX_L$ = *total self-impedance in loop* 1 with all other loops open-circuited

$\mathbf{Z}_{22} = j(X_L - X_C)$ = total self-impedance in loop 2 with all other loops open-circuited

$\mathbf{Z}_{12} = -jX_L$ = mutual impedance between loops 1 and 2

$\mathbf{Z}_{21} = -jX_L$ = mutual impedance between loops 2 and 1

For a bilateral network (one which contains elements *that transmit energy equally in either direction in a circuit*),

$$\mathbf{Z}_{12} = \mathbf{Z}_{21} \quad (14)$$

$\mathbf{Z}_{12}$ is equal to $-jX_L$ in this particular problem because I_1 and I_2 were chosen opposite through the mutual branch L.

L-loop network By induction, the standard form for the loop equations in an L-loop network may be written

$$\begin{aligned}
\mathbf{E}_1 &= \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 + \mathbf{Z}_{13}\mathbf{I}_3 + \quad + \mathbf{Z}_{1L}\mathbf{I}_L \\
\mathbf{E}_2 &= \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 + \mathbf{Z}_{23}\mathbf{I}_3 + \quad + \mathbf{Z}_{2L}\mathbf{I}_L \\
\mathbf{E}_3 &= \mathbf{Z}_{31}\mathbf{I}_1 + \mathbf{Z}_{32}\mathbf{I}_2 + \mathbf{Z}_{33}\mathbf{I}_3 + \quad + \mathbf{Z}_{3L}\mathbf{I}_L \quad (15) \\
&\cdots \\
\mathbf{E}_L &= \mathbf{Z}_{L1}\mathbf{I}_1 + \mathbf{Z}_{L2}\mathbf{I}_2 + \mathbf{Z}_{L3}\mathbf{I}_3 + \quad + \mathbf{Z}_{LL}\mathbf{I}_L
\end{aligned}$$

where $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_L$ are the total emfs in each of the loops; $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_L$ are the loop currents; $\mathbf{Z}_{12}, \ldots, \mathbf{Z}_{1L}, \ldots, \mathbf{Z}_{L1}, \mathbf{Z}_{L2}, \ldots$ are the mutual impedances. For bilateral networks

$$\mathbf{Z}_{pq} = \mathbf{Z}_{qp} \quad (16)$$

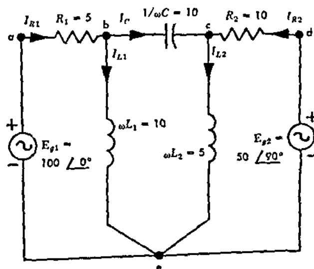


Fig. 4

where p and q represent adjacent loops with mutual impedance.

Node method. Before illustrating the node method of organizing the algebraic work for finding the branch currents, it is desirable to indicate how the branch currents may be expressed in terms of node voltages (see Fig. 4).

The junction points a, b, c, d, and e are called nodes. Nodes with more than two branches, such as b, c, and e, are called major nodes. If e is chosen *arbitrarily as the reference node for voltages,* then $\mathbf{V}_a = \mathbf{V}_{ae}$ is the node voltage drop (+ to −) from a to e; similarly for the node voltages $\mathbf{V}_b = \mathbf{V}_{be}$, $\mathbf{V}_c = \mathbf{V}_{ce}$, and $\mathbf{V}_d = \mathbf{V}_{de}$. The branch currents in terms of node voltages are

$$\mathbf{I}_{R1} = \frac{\mathbf{V}_a - \mathbf{V}_b}{R_1} \qquad \mathbf{I}_{L1} = \frac{\mathbf{V}_b}{j\omega L_1} \qquad \mathbf{I}_C = \frac{\mathbf{V}_b - \mathbf{V}_c}{-j\dfrac{1}{\omega C}} \quad (17)$$

$$\mathbf{I}_{L2} = \frac{\mathbf{V}_c}{j\omega L_2} \qquad \mathbf{I}_{R2} = \frac{\mathbf{V}_d - \mathbf{V}_c}{R_2}$$

The current or node equation at major node b is

$$\mathbf{I}_{R1} = \mathbf{I}_{L1} + \mathbf{I}_C \quad (18)$$

or

$$\frac{\mathbf{V}_a - \mathbf{V}_b}{R_1} = \frac{\mathbf{V}_b}{j\omega L_1} + \frac{\mathbf{V}_b - \mathbf{V}_c}{-j\dfrac{1}{\omega C}} \quad (19)$$

The node equation at major node c is

$$\mathbf{I}_C + \mathbf{I}_{R2} = \mathbf{I}_{L2} \quad (20)$$

or

$$\frac{\mathbf{V}_b - \mathbf{V}_c}{-j\dfrac{1}{\omega C}} + \frac{\mathbf{V}_d - \mathbf{V}_c}{R_2} = \frac{\mathbf{V}_c}{j\omega L_2} \quad (21)$$

The node voltages are the unknowns in the node equations (19) and (21).

The node method consists of writing only the node equations with branch currents expressed in terms of node voltages. The node voltages now are the unknowns in the node equations. Solving the node equations simultaneously will give values for the node voltages which can be substituted in the expressions for the branch currents, such as those in Eq. (17). The node method may be outlined in the following steps:

Step 1. Determine the number of independent node equations by the relation $n = (n_m - 1)$

Step 2 Write the node equations at the major nodes in the form such as (19) or (21).

Step 3 Substitute known values for as many node voltages as possible The number of remaining unknown node voltages should be equal to n in step 1

Step 4. Solve the node equations simultaneously for the node voltages.

Step 5. Substitute in Eq (17) and obtain the branch currents.

The circuit of Fig. 4 is solved as follows.

Step 1. The number of major nodes in $n_m = 3$ and the number of independent node equations is $n = (n_m - 1) = 2$. That is, only two independent

neural tube the crest cells migrate into the body and give rise to parts of the neural system, the main part of the visceral cranium, the mesenchyme, chromaffin cells, and to pigment cells The true nature of the neural crest was revealed only recently because this primary organ has a temporary existence; its cells and derivatives are difficult to analyze when dispersed throughout the body. The fact that mesenchyme arises from this ectodermal organ was directly contrary to the doctrine of the specificity of the germ layers *See* GERM LAYERS

Neural crest no doubt exists, with similar qualities, in all vertebrate groups, including the cyclostomes. It has been most thoroughly studied in *amphibians*

Crest cells. The dorsal ectoderm of a vertebrate gastrula forms the pear-shaped neural plate which is surrounded by the neural ridge (Fig 1). When the neural plate rolls up to form the neural tube, the ridges from each side meet and fuse, temporarily forming a wedge shaped cell mass in the dorsal midline of the prospective brain and spinal cord. In Fig 1 the head part of the ridge of an urodele neurula is divided into eight zones Zones 1 and 2 may be called the transverse ridge. Zone 8 and the posterior parts are the trunk ridge. The neural crest cells are situated in the ridge but do not occupy the whole ridge For example, the posterior slope of the transverse ridge consists of material for the forebrain, and the anterior slope forms epidermis (border broken dotted line on left side, Fig 1) Recent investigations have confirmed that probably no crest cells emanate from the transverse ridge (C Jacobson, 1959, P. Nieuwkoop,

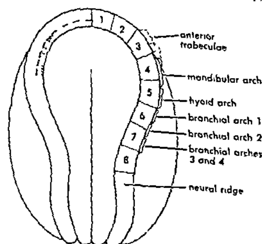


Fig 1 Diagram of open neural-plate stage of a urodele, *Amblystoma mexicanum* Left side, epidermal line of coalescence (*broken line*), line of coalescence of the brain (dotted line), presumptive ectomesenchyme (hatched area) Right side, position of presumptive anterior trabeculae, mandibular arch, hyoid arch, and branchial arches in relation to zones 1–8. (From S Hörstadius, *The Neural Crest, Its Properties and Derivatives in the Light of Experimental Research*, Oxford, 1950)

I. Oikawa, and J. Boddingius, 1958) Concerning the main parts of the neural ridge it is still doubtful whether the crest is located only in the peripheral part of the thick plate or in the adjacent part of the thinner ectoderm as well. At least in zones 1-4 of the *Amblystoma* neurula the line of coalescence of the epidermis coincides with the midline of the ridge, the crest material thus forming the proximal slope of the ridge (Fig 1, hatched area). The exact position of the material in the posterior parts and in other vertebrate groups is not known.

The crest cells do not remain long in the dorsal midline of the neural tube They migrate in a lateral direction between the tube and epidermis, first as rather coherent sheets, then in groups, strands, or as single cells, both proximal and distal to the somites. Because they are similar to the mesenchymal cells originating from the mesoderm, their fate has been ascertained only by vital staining, extirpation, or transplantation experiments in many cases. One method has been to replace a part of the ridge by a corresponding piece from another species with cell nuclei of different size which are thus recognizable in sections. Most extirpations and transplantations [illegible]

[illegible] operated side, and extirpations must also include rather long pieces of the ridge because the gap may be filled by cells from the crest anterior and posterior to the excision Transplantation of neural crest to another part of the body as well as explantation have been used in order to determine what it can give rise to in new surroundings. In spite of all efforts, however, the problem of the fate of the neural crest is still not solved in all its details. *See* FATE MAPS, EMBRYONIC.

Contributions to the neural system. Parts of the migrating bilateral cell masses settle in clusters and give rise to the spinal ganglia. The segmental arrangement of the dorsal nerve roots and the ganglia is not intrinsic to the cells but is determined by the segmental arrangement of the somites. There seems to be no doubt at the present time that the spinal ganglia are formed entirely of neural crest cells. Concerning the origin of the cranial ganglia, there has been much controversy. Placodes of the lateral epidermis of the head are one source of their cells; the other is the neural crest, namely for the sensory components of the ganglia of the nerves V, VII, IX, and X, although to different degrees in different groups of animals. *See* NERVOUS SYSTEM.

According to both histological observations and experiments, the sympathetic trunk seems to be derived not only from the crest but also from the ventral part of the spinal cord, however, crest contribution has been denied by some. Excision experiments on chicks support crest origin (G Nawar, 1956) Cells from the sympathetic trunks migrate farther away to form the neurons of the prevertebral ganglia (celiac and mesenteric).

with Z_L removed and whose series impedance is the impedance of the network looking back from terminals *a-b* into the network with all generators replaced by their internal impedances.

Thévenin's theorem is useful in complicated networks where Z_L is being varied, as in the case of maximum power transfer. It should be understood that the Thévenin generator is equivalent only so far as the current through Z_L is concerned. *See* THÉVENIN'S THEOREM (ELECTRIC NETWORKS).

Norton's theorem. The current in any impedance Z_L between terminals *a-b* of a linear bilateral network containing generators of the same frequency is equal to the current in the same Z_L when it is connected to a current generator whose generated current I_{SC} is that current which flows through terminals *a-b* when these terminals are short-circuited and whose shunt impedance is the impedance of the network looking back from terminals *a-b* into the network with all generators replaced by their internal impedances.

In many respects, Norton's theorem is similar to Thévenin's theorem. The Norton generator is equivalent only so far as the current through Z_L is concerned Thévenin's theorem and Norton's theorem indicate that a Thévenin generator may be transformed into a Norton generator; and a Norton generator into a Thévenin generator.

Maximum power transfer theorem. A linear bilateral network containing generators of the same frequency can be represented, from Thévenin's theorem, by a voltage generator such as that to the left of *a-b* in Fig. 7. The relations between Z_L and Z_{ab} for maximum power transfer to Z_L given here are for the conditions that Z_{ab} is a fixed impedance, Z_L is a variable impedance, and E_o is a constant voltage.

When $Z_{ab} = R_{ab}$ and $Z_L = R_L$, then for maximum power transfer,

$$R_L = R_{ab}$$

When $Z_{ab} = R_{ab} \pm jX_{ab}$ and $Z_L = R_L \pm jX_L$, then

$$Z_L = Z^*_{ab}$$

where Z^*_{ab} is the conjugate of Z_{ab} $(= R_{ab} \pm jX_{ab})$.

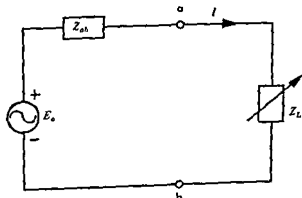
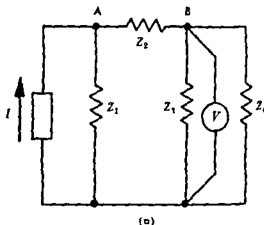


Fig. 7.



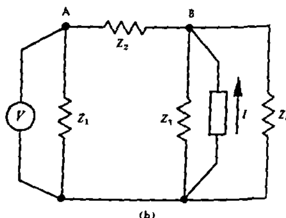


Fig 8 (a–b).

When $Z_{ab} = Z_{ab}/\underline{\theta_{ab}}$ and $Z_L = Z_L/\underline{\theta_L}$ where θ_L is fixed, then

$$Z_L = Z_{ab}$$

Reciprocity theorem. In any passive, linear, bilateral network, the current I, flowing in any branch *a* as a result of an emf E_g applied in a second branch *b*, is equal to the current that would be produced in the second branch *b* if E_g were transferred to the first branch *a*, provided that either equal impedances or no impedances are interchanged in this transfer of the source of emf E_g.

This theorem shows that any linear, bilateral network transmits power with equal effectiveness in both directions, if the generator and the load being interchanged have the same impedance Also, this theorem says nothing about the network currents or voltages not involved in the reciprocity

The so-called dual of this theorem may be stated as follows. if a current applied at node A pro-

PLE OF. [K.Y.T.]

Neural crest

A strip of ectodermal material in the early vertebrate embryo inserted between the prospective neural plate and epidermis. After closure of the

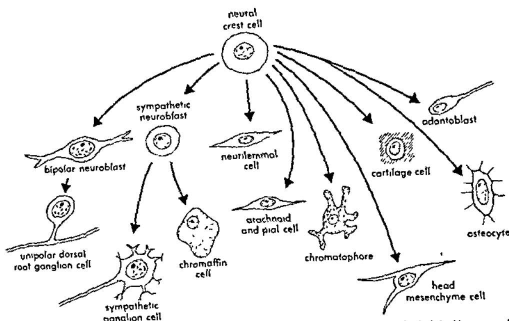


Fig 3 A schematic representation of the possible developmental fates of a neural crest cell (*From W. J. Hamilton, J D Boyd, and H. W Mossman, Human Embryology, 2d ed, Williams and Wilkins, 1952*)

prospective pigment cells in amphibians may remain dormant during larval development and become activated at metamorphosis The activation of the propigment is evidently brought about by an oxidase The pigment pattern depends both on properties intrinsic in the propigment cells and on the surrounding tissues Also when transplanted to a host of very distant relationship as from anuran to urodele, the single cells differentiate in size shape and color as they would have done in the donor However their distribution may vary

In reciprocal exchange of neural crest between some urodele species the *pattern in the respective regions* becomes that of the donor, but in less related forms influence of the host is more or less *recognizable* Grafting of epidermis from another species has hardly any effect In some urodele species there is an intrinsic tendency in the crest cells to aggregate at certain levels, for example, along the dorsal border of the somites. Melanoblasts invade developing feathers, producing their specific color and pattern They also invade growing hair giving pigment granules to it But in some cases, both in the bird and *mammal* the color is a differentiation dependent upon the epidermis The black and white stripes in the Barred Plymouth Rock feather pattern are probably the result of an inhibitory action from each black stripe on the melanoblasts of the prospective white stripe during feather growth *See* CHROMATOPHORE, EMBRYONIC INDUCTION, EYE

It is not known what factor influences the pathway of the migrating pigment cells. Grafted cells may move against or perpendicular to their normal path

[S H]

Bibliography. J P Hill and K M Watson, The early development of the brain in marsupials, *J. Anat.*, 92(4):493–497, 1958; S. O. Horstadius, *The Neural Crest, Its Properties and Derivatives in the Light of Experimental Research*, 1950; C. O Jacobson, The localisation of the presumptive cerebral regions in the neural plate of the axolotl larva, *J. Embryol. and Exptl. Morphol.*, 7(1):1–21, 1959, G. Nawar, Experimental analysis of the origin of the autonomic ganglia in the chick embryo, *Am. J Anat*, 99(3):473–505, 1956; P. D. Nieuwkoop, I Oikawa, and J. Boddingius, The anterior transverse neural fold in amphibians, *Arch néerl. zool.*, suppl. 1, 13:167–184, 1958; M E Rawles, Origin of melanophores and their role in development of the color patterns in vertebrates, *Physiol. Rev.*, 28: 383–408, 1948, T Seno and P D Nieuwkoop, The autonomous and dependent differentiations of the neural crest in amphibians, *Koninkl. Ned. Akad. Wetenschap Proc* Ser C, 61:489–498, 1958; D Starck, *Embryologie; Ein Lehrbuch auf allgemein biologischer Grundlage*, 1955.

Neuron

The developmental, structural, and functional unit of the *nervous system*. It is composed of a nerve cell body [illegible] a single process, is the axon or axis cylinder which carries the impulse away from the cell body. *Nerve impulses* pass from *one neuron* to another at discontinuous junctions where the nerve processes of *one neuron* impinge upon another nerve cell or its processes. *The nervous system* contains an enormous number of neurons of specific shapes, sizes, and location for each anatomic or functional group. *See* NERVOUS SYSTEM

[I C ST.]

Whether crest cells also contribute to the parasympathetic system is disputed.

The cellular sheath of Schwann enclosing peripheral nerves is probably formed mainly of crest cells, but it seems that in later stages neurilemmal cells may also emerge from the neural tube by way of ventral roots. However, heteroplastic transplantations with nuclei of different size lead to the belief that the lateral walls of the spinal cord are their sole source. The majority of the leptomeninx (pia and arachnoidea) cells are derived from the neural crest whereas the pachymeninx (dura) is composed chiefly of endomesodermal cells. For differentiation of crest cells into cells of Schwann and meninges the presence of neural material seems indispensable (T. Seno and P. Nieuwkoop, 1958).

Chromaffin cells. The chromaffin cells of the suprarenal organs of lower vertebrates as well as the medullary zone of the suprarenal gland in mammals, and also those of the paraganglia, that is, the aortic bodies of Zuckerkandl and the carotid bodies, are all considered to be formed by cells emanating from the sympathetic system, thus probably mainly from the crest See ADRENAL GLAND; AORTIC BODY.

Mesenchyme and skeleton. The statement by J. B. Platt in the 1880s that mesenchyme could arise from ectoderm (mesectoderm, ectomesoderm, or ectomesenchyme) raised a vivid discussion for several decades. Experiments on amphibian larvae have confirmed that of the chondrocranium, the anterior half of the trabeculae, all visceral arches, and the first basibranchial are of neural crest origin, whereas the main part of the neurocranium and the second basibranchial are endomesodermal (Fig. 2). In addition membrane bones, namely the tooth-bearing premaxilla, vomeropalatine, dentary, and splenial, as well as the odontoblasts and dentine, are formed by cells emanating from the neural crest Moreover, the ectomesenchyme of the teeth induces the formation of enamel organs. Crest origin of the mandibular arches has been described in marsupials by J. Hill and K. Watson in 1958. See SKELETAL SYSTEM.

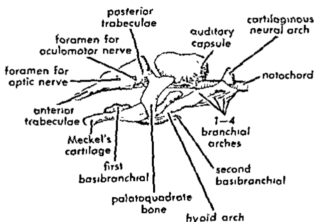


Fig 2. Chondrocranium from left of larva of *Amblystoma mexicanum*. Cartilage of endomesodermal origin dotted. (Modified from S. Hörstadius, *The Neural Crest; Its Properties and Derivatives in the Light of Experimental Research*, Oxford, 1950)

The transverse ridge, zones 1 and 2 in Fig. 1, has no cartilage-forming capacity. The trabecular material is situated in zone 3 which cannot partake in formation of the visceral arches. The mandibular ectomesenchyme is located in zone 4 and perhaps also in the adjacent part of zone 3, which cannot form trabeculae but is capable of forming other visceral arches, although there is no fusion with the basibranchials; the material of the hyoid arch and the gill arches in zones 5–8 shows such capacity to fuse. In zone 5 the trunk crest begins; it has no faculty to form cartilage but differentiates to form spinal and sympathetic ganglia as well as chromaffin, pigment, and other mesenchymal cells. It has not yet been determined that crest cells in the neural ridge of the head form cartilage. No cartilage is formed when they are transplanted to the trunk unless they are activated by certain tissues such as the pharyngeal endoderm and to some extent the gastric endoderm, gill-region endomesoderm, chorda, and wounds in somites.

To what extent corium cells are of crest origin is not clearly established. Head mesenchyme is also formed by the crest. In urodele larvae no dorsal fin is formed, nor do the larval gills develop normally in absence of the crest.

Pigment cells. With the exception of the retinal pigment, which is formed in situ, probably all chromatophores in vertebrates are derived from the neural crest. Such an origin has been observed in fishes, amphibians, birds, and mammals. Pigment cells are found in both dermis and epidermis, feathers, hair, the perineural and perivascular layers, the coelomic wall, and the choroid and iris of the eye The origin of pigment cells was first traced in this century. Because the pigment is formed rather late in embryonic development, the presumptive pigment cells cannot be distinguished from other mesenchyme cells in early stages There are brown to black melanophores, yellow lipophores (xanthophores), reddish erythrophores (allophores), and guanophores and iridocytes with resplendent crystals. Bilateral extirpation of sufficiently large pieces of neural crest results in unpigmented regions of the body Following extirpation of all the neural ridges, urodele larvae subsequently obtain pigment cells emerging from the brain and the tail-bud part of the neural plate. Normally the reserve in these regions is probably inhibited by the pigment cells from the crest Explants of neural crest may give pigment cells in vitro. Implantation of neural crest into a larva of another species has produced pigment of the type of the donor in amphibians, birds, and mammals. The development of pigment in the chromatophore is no self-differentiation but is dependent upon factors in surrounding cells. For example, chromatophores of the black axolotl produce no pigment in white specimens, whereas chromatophores from the white race are activated and form melanin granules in the skin of a black specimen.

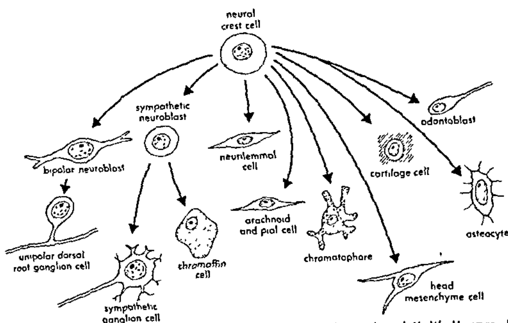


Fig 3 A schematic representation of the possible developmental fates of a neural crest cell (From W J Hamilton, J D Boyd, and H. W. Mossman, *Human Embryology*, 2d ed , Williams and Wilkins, 1952)

prospective pigment cells in amphibians may remain dormant during larval development and become activated at metamorphosis The activation of the propigment is evidently brought about by an oxidase. The pigment pattern depends both on properties intrinsic in the propigment cells and on the surrounding tissues. Also when transplanted to a host of very distant relationship as from anuran to urodele, the single cells differentiate in size shape, and color as they would have done in the donor However their distribution may vary

In reciprocal exchange of neural crest between some urodele species the pattern in the respective regions becomes that of the donor, but in less related forms influence of the host is more or less recognizable Grafting of epidermis from another species has hardly any effect In some urodele species there is an intrinsic tendency in the crest cells to aggregate at certain levels for example, along the dorsal border of the somites Melanoblasts invade developing feathers, producing their specific color and pattern They also invade growing hair giving pigment granules to it But in some cases, both in the bird and mammal, the color is a differentiation dependent upon the epidermis The black and white stripes in the Barred Plymouth Rock feather pattern are probably the result of an inhibitory action from each black stripe on the melanoblasts of the prospective white stripe during feather growth See CHROMATOPHORE, EMBRYONIC INDUCTION; EYE

It is not known what factor influences the pathway of the migrating pigment cells Grafted cells may move against or perpendicular to their normal path [S H.]

Bibliography. J P. Hill and K. M. Watson, The early development of the brain in marsupials, *J Anat*., 92(4):493–497, 1958; S. O Horstadius, *The Neural Crest, Its Properties and Derivatives in the Light of Experimental Research*, 1950, C. O. Jacobson, The localisation of the presumptive cerebral regions in the neural plate of the axolotl larva, *J Embryol and Exptl Morphol*., 7(1):1–21, 1959; G Nawar, Experimental analysis of the origin of the autonomic ganglia in the chick embryo, *Am. J Anat*., 99(3):473–505, 1956, P D. Nieuwkoop, I Oikawa, and J Boddingius, The anterior transverse neural fold in amphibians, *Arch. néerl. zool*., suppl 1, 13 167–184, 1958, M E Rawles, Origin of melanophores and their role in development of the c[illegible] 383–[illegible] autor[illegible] neural crest in amphibians, *Koninkl. Ned. Akad. Wetenschap. Proc* Ser. C, 61:489–498, 1958; D. Starck, *Embryologie, Ein Lehrbuch auf allgemein biologischer Grundlage*, 1955.

Neuron

The developmental, structural, and functional unit of the nervous system It is composed of a nerve cell body, containing the nucleus, and two types of processes, or nerve fiber The first type, the dendrites, carry the unidirectional nerve impulse to the cell body; the second type, ordinarily a single process, is the axon or axis cylinder which carries the impulse away from the cell body. Nerve impulses pass from one neuron to another at discontinuous junctions where the nerve processes of one neuron impinge upon another nerve cell or its processes. The nervous system contains an enormous number of neurons of specific shapes, sizes, and location for each anatomic or functional group See NERVOUS SYSTEM. [E G.ST]

Neurophysiology

The sum of the facts and generalizations which describe the activities of the nervous system and its parts.

An organism's nervous system is its principal means of rapidly coordinating its activities to meet the demands of both the external and internal environment. The organism's reactions to the external world take place by means of striped (striated) muscles which are activated by nerves coming directly from the spinal cord. The physical well-being of an organism (its internal environment) is regulated by smooth muscles and glands whose stimulation comes from ganglia which are in turn activated by nerves from the spinal cord. The nerves to all muscles and glands leave the spinal cord on the side nearest the intestines (ventral), whereas the nerves from sense organs enter through roots close to the back (dorsal). The spinal cord together with its sensory and motor nerves is capable of regulating blood flow and mediating such behavior as scratching, the maintenance of posture, standing, and elements of walking Other forms of activity, however, depend on the integrity of some portion of the brain. *See* SPINAL CORD.

The brain may be divided into the hindbrain, midbrain, and forebrain The forebrain is further divided into the between-brain and endbrain Breathing and vomiting depend upon structures in the ventral part of the hindbrain, the part closest to the spinal cord. The maintenance of posture and position, adjustment of gait, and the detection of differences between spatial cues are activities mediated in part by the cerebellum, the dorsal part of the hindbrain. In the ventral part of the midbrain there are structures needed for maintenance of wakefulness, which are also possibly basic to attention and consciousness Temperature regulation, food and water intake, and the production of some hormones are mediated by a structure in the lower part of the between-brain called the hypothalamus. The hypothalamus and some anatomically related areas are sites which a rat or a monkey will repeatedly stimulate electrically, if the stimulation is under the animal's control. *See* BRAIN; ELECTROENCEPHALOGRAPHY.

The cerebral cortex, a part of the endbrain, is more developed in man and primates than in other vertebrates and is considered to mediate the most advanced forms of behavior. The limited activities of an animal deprived of cortex confirm this. The cortex is the end station of most sensory pathways and the place of origin of a primary motor tract. The functions of the sensory cortex are exemplified by the auditory cortex, which mediates localization of sound in space and the discrimination of sound patterns. In addition to the motor cortex and sensory cortex there are large areas of cortex whose functions are not accurately known. *See* PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL.

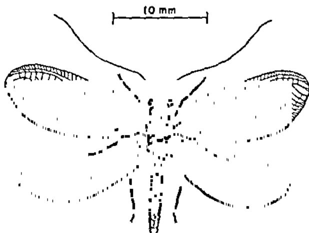
[J.P.F.]

Neuroptera

An order of rather fragile insects with complete metamorphosis and chewing mouthparts. They are small to rather large-sized, soft-bodied organisms. Common names for members of this group are the lacewings, ant lions, dobson flies, and raphidids. The order is rather small, consisting of some 40 families, approximately 350 genera, and about 4000 species. It is world-wide in distribution, being found in practically all localities where plant and animal life exists. The oldest member of the order dates from the Upper Permian of Belmont, Australia.

Members of the order are mostly terrestrial and are considered beneficial because of their carnivorous habits. The larvae and adults of some families are voracious predators upon many types of small insects and mites.

The adults have long antennae and four similar, large, membranous wings with many cross veins.



Chrysopidae, adult European pearly lacewing, Chrysopa perla (L.) (From E O. Essig, College Entomology, Macmillan, 1942)

Most members, in repose, hold their wings rooflike over the back. Cerci are absent. Larvae are thysanuriform, robust, or spindle-shaped, with biting-type mouthparts modified for piercing and sucking.

Pupation occurs in silken cocoons, and the adults escape through a hole made in the cocoon by cutting out a disk with their well-developed mandibles *See* INSECTA; PTERYGOTA, *see also* ANT LION; HELLGRAMITE. [E O E]

Bibliography: E. O. Essig, *College Entomology*, 1942.

Neurosis

A term generally used to refer to disabling psychological states that impair normal mental functioning but at the same time do not totally disrupt and disorganize the person's grasp of the realities with which he must cope. The processes that lead to neurosis are regarded as aggravated reactions

to life problems, the aggravated nature of the reaction being attributable to internal conflicts and inappropriate defenses that make it extremely difficult for the person to cope with either external demands or the demands of his own needs

The symptoms expressed in the neurosis may often be desperate attempts to deal with these problems—attempts that do not or cannot resolve the difficulties. In the sense that neurotic symptomatology is consistent with the pattern of the person's general mode of functioning, neurosis may be conceived of as reflecting a continuation of the psychodynamic pattern responsible for character formation in the individual (see PERSONALITY THEORY) Neurosis is not a disease entity but an exaggerated continuation of mental functioning in the face of internal and external trouble and stress In short, the kinds of behaviors that are found in extreme form in the neurotic are found in less extreme form in the normal person

R W White has stated that "the core of a neurosis lies at the point where anxiety has blocked or distorted the learning process so that new learning essential to adjustment cannot take place" When the person is confronted by deeply threatening demands in the sense of their being unacceptable to him, anxiety results Severe anxiety, marked by the absence of any discernible, appropriate cause, has the effect of producing emergency reaction whose object is to bring the disrupting state to an end as rapidly as possible. This reaction may take various forms, and the form that it does take defines the symptomatology of the neurosis Because anxiety makes it virtually impossible for the person to learn directly how to cope with his difficulty, the reaction that develops becomes self-defeatingly maladaptive Because the neurotic reaction has at least the effect of reducing the terrors of anxiety, it has a gain feature to it that leads the person to cling to his symptoms, and indeed, to resist their removal This fixity of the neurosis is further reinforced by the secondary gains obtained through the development of a sick role with attendant sympathy and special care

Four general patterns of neurotic reaction, corresponding to different, though overlapping, modes of defense may be distinguished the hysteric, the phobic and the counterphobic, the neuroses involving somatization, and the obsessional compulsive These neurotic patterns may be either precipitated by chronic conditions or seemingly traumatic in origin

The major forms of therapy used in dealing with neurotic problems are psychological rather than physiological in nature, although mild tranquilizers and sedatives are occasionally used to alleviate disruptive anxiety states At the base of most psychotherapy, usually strongly influenced by psychoanalytic theory and practice, is the formation of a relationship between the therapist and patient that makes it possible for the latter to verbalize freely the nature of his difficulties and, with the help of the therapist, to search for some insight as to the origins and meaning of his difficulties. The process of psychotherapy tends to be slow because of the operation of defenses on the part of the patient and because there are secondary gains that lead the patient, as Freud once said, to "cling to the neurosis" in spite of his conscious and often ardently expressed wish to get well. *See* ABNORMAL BEHAVIOR; HYSTERIA; OBSESSIVE COMPULSIVE REACTION; PHOBIC REACTION; SOMATIZATION.

[I.S.B.; W.MS.]

Bibliography: O Fenichel, *The Psychoanalytic Theory of the Neuroses*, 1945; S. Freud, *Collected Papers*, vols. 1-5, 1953; D. Henderson and R. D. Gillespie, *Textbook of Psychiatry for Students and Practitioners*, 1927; J. McV. Hunt (ed), *Personality and the Behavior Disorders*, 1944; R. W. White, *The Abnormal Personality*, 2d ed , 1956.

Neurulation

The process by which the vertebrate neural tube is formed The primordium of the central nervous system is the neural plate, which arises at the close of gastrulation by inductive action of the chordamesoderm on the overlying ectoderm. The axial mesodermal substratum causes the neural ectoderm to thicken into a distinct plate across the dorsal midline and influences both its size and shape Its shieldlike appearance, broader anteriorly and narrower posteriorly, presages the future areas of brain and spinal cord, respectively.

The lateral edges of the neural plate then rise as neural folds which meet first at the level of the future midbrain, above the dorsal midline, then fuse anteriorly and posteriorly to form the neural tube The body ectoderm becomes confluent above the closing neural tube and separates from it Upon closure, the cells (known as neural crest cells) which occupied the crest of the neural folds leave the roof of the tube and migrate through the mesenchyme to all parts of the embryo, forming diverse structures

The neural tube thus formed gives rise to the brain and about half of the spinal cord The remainder of the neural tube is added by the tail bud, which proliferates a solid nerve cord that secondarily hollows into a tube.

The closure of the neural tube *confines* secretions from its inner wall, which dilate the central canal by their *turgor* and help to expand [illegible]

thickness or thinness of the neural wall. Adjacent to the notochord there are few mitoses and the floor of the tube remains thin, while in the somite region the lateral walls are actively mitotic and become thick The contour is also passively molded by adjacent structures [illegible]

[illegible] *neural cells* differentiating into neurons.

Neuromeres are temporary constrictions in the hindbrain resulting from localized growth processes. They were once thought to have evolutionary significance as vestiges of metamerism, but the plasticity of the neural tube with respect to its mesodermal surroundings makes this interpretation doubtful.

Neural folds never form in teleost fish. The neural plate concentrates into a solid neural keel which then hollows out secondarily as in the tail bud of other vertebrates See EMBRYOLOGY; EMBRYONIC INDUCTION; GASTRULATION; NERVOUS SYSTEM; NEURAL CREST [H.L.H.]

Neutralization

In chemical literature, processes in which the acidity or basicity of a solution is destroyed (neutralized) by the reaction of hydrogen ion with hydroxide ion to produce water.

When a solution containing exactly 1 mole of hydrochloric acid in aqueous solution is mixed with an aqueous solution containing exactly 1 mole of sodium hydroxide, the result is a solution containing 1 mole of the salt sodium chloride and the water which was present in both of the solutions before mixing plus 1 mole produced by the reaction. The resulting solution is said to be neutral because the hydrogen ion and hydroxide ion concentrations are equal.

When, on the other hand, 1 mole of acetic acid is mixed with exactly 1 mole of sodium hydroxide, the final solution is slightly basic because the salt produced is hydrolyzed (*see* HYDROLYSIS) In this example neutralization does not produce a perfectly neutral solution A common analytical procedure is the determination of the amount of an acid or base present by neutralization of some unknown amount of that substance with a carefully measured volume (or weight) of a standardized solution (one containing a known amount of reagent in a specified volume or weight) The end of such a titration is usually indicated by a very rapid change in the pH. A rapid change in pH can be detected accurately either by electrical methods or by indicators (chemical substances which change color when pH is altered within a fairly narrow range)

The word *neutralize* is sometimes used in the general sense of *nullify* The effects of many chemicals or of change of conditions may be nullified by appropriate action of the experimenter, such as addition of other chemicals or appropriate change of conditions. *See* ACID AND BASE; TITRATION.

[T F Y]

Bibliography· W. C Pierce, E L Haenisch, and D. T. Sawyer, *Quantitative Analysis*, 1958

Neutralization reaction (antibody)

A procedure in which the chemical or biological activity of a reagent or a living organism is inhibited, usually by a specific neutralizing antibody. As an example, the lethal or the dermonecrotic actions of diphtheria toxin on animals may be completely neutralized by an equivalent amount of diphtheria antitoxin—an antibody produced in animals or in humans after contact with diphtheria toxin or toxoid. Lesser amounts of antitoxin provide intermediate degrees of inhibition. These facts provide the basis for the Schick test for susceptibility to diphtheria. Tetanus and botulinus toxins may be similarly inhibited by their specific antitoxins. In contrast, the typical toxins of dysentery and other gram-negative bacteria are only slightly neutralized, even by large excesses of antibody. Antibodies to bacterial, snake venom, and other enzyme preparations regularly precipitate them from solution so that the supernates are devoid of enzyme activity; however, the neutralization of activity in the precipitate may range from complete to negligible See BACILLARY DYSENTERY; BOTULISM; IMMUNOLOGY, NEUTRALIZING ANTIBODY; SEROLOGY; TETANUS

Infection of a host by a living bacterium, virus, or other microorganism may also be inhibited or mitigated by the corresponding antibodies, and such neutralization tests are used in the diagnostic examination of sera or of infective agents recovered from such infections as poliomyelitis and yellow fever In some instances, the antibody may be injected into a test animal before, or occasionally shortly after, challenge with the living agent In other instances, the neutralization of the microbial infectivity by the antibody is permitted to take place in the test tube, and its degree determined by subsequent injection of the mixture into an appropriate test animal [H P.T.]

Bibliography. W. C Boyd, *Fundamentals of Immunology*, 3d ed., 1956; T M. Rivers and F. L. Horsfall, Jr. (eds), *Viral and Rickettsial Infections of Man*, 3d ed , 1959

Neutralizing antibody

An antibody that reduces or abolishes some biological activity of a soluble antigen or of a living microorganism Thus, diphtheria antitoxin is a neutralizing antibody that, in adequate amounts, abolishes the pathological effects of diphtheria toxin in animals.

Analogous neutralizing effects of antibodies can be demonstrated for the lytic effects of many lysins and, most important, for the pathogenic effects of viruses and the rickettsiae Since the latter are complex bodies containing multiple antigens not all of the resulting antibodies need have neutralizing activities, although they may display a variety of other serological properties. Antibodies to enzymes constitute a special case, in all instances, they precipitate their corresponding enzyme, but the degree of neutralization may range from 0 to 100%. Antibodies to the endotoxins of the gram-negative bacteria regularly neutralize their toxicity only to a low degree. *See* ANTIGEN; ANTITOXIN; DIPHTHERIA; LYSIN; NEUTRALIZATION REACTION (ANTIBODY); RICKETTSIOSES; TOXIN, BACTERIAL. [H.P T.]

Neutrino

A neutral particle having zero rest mass and spin ½(h/2π), where h is Planck's constant. It is emitted in β decay and other weak decays. The existence of the neutrino was originally postulated to account for the energy, linear momentum, and angular momentum which is missing from the observed particles (the residual nucleus and the electron) resulting from a β-decay. The neutrino, denoted by the Greek letter ν, is a lepton which obeys Fermi-Dirac statistics; its only known interaction is the β-interaction, which is so weak that the direct observation of the neutrino is very difficult. *See* ELEMENTARY PARTICLE, LEPTON; RADIOACTIVITY.

The antineutrino ($\bar{\nu}$) is distinct from the neutrino, it is an antilepton. A neutrino is emitted in positron β-decay (for example, $A^{37} \rightarrow Cl^{37} + e^+ + \nu$); an antineutrino is emitted in electron β-decay (for example, $n \rightarrow p + e^- + \bar{\nu}$).

The neutrino is massless, like the photon, and travels at the speed of light. Its masslessness also implies that in a pure spin state its spin can be directed only along its motion or opposite to its motion. In the former case the neutrino is said to have a right-hand polarization, or briefly, to be right handed. (If the extended thumb of a right hand is taken to point in the direction of motion of the neutrino, the fingers, partially closed toward the palm, indicate the sense of spin of the right-handed neutrino.) Conversely, a neutrino whose spin is directed oppositely to its motion is polarized left-handed. Because of its masslessness, the handedness of a neutrino is an intrinsic property, like its charge or rest mass. (The apparent handedness of a massive particle, however, depends on the observer. If the observer travels faster than the particle, he will see its momentum, and therefore its handedness, reversed. A synonym of handedness is helicity; positive helicity = right handed.)

Two-component theory. It is found by experiment that only left handed neutrinos exist, and that the antineutrino is right-handed. Thus the neutrino field has only two possible quanta (left handed, ν, and right handed, $\bar{\nu}$), rather than the four ordinarily possible for a spin ½ particle (right and left-handed particle and antiparticle). Therefore the neutrino field need only have two components, rather than the four specified by the Dirac equation (see QUANTUM FIELD THEORY). Th[illegible]
b[illegible]
b[illegible] parity would then be violated. For upon inv[illegible] becomes righ[illegible] inverted wor[illegible] this fact in [illegible] theory of parity nonconservation in weak interactions, conjecturing that only neutrinos of one handedness actually exist. *See* PARITY (QUANTUM MECHANICS); SYMMETRY LAWS (PHYSICS).

Inverse beta decay. This provides the direct evidence for the neutrino and was observed by F. Reines and C. L. Cowan in spite of the exceedingly small interaction cross section of the neutrino, ~10^{-44} cm². A nuclear reactor is a strong source of antineutrinos, which may undergo the reaction $\bar{\nu} + p \rightarrow n + e^+$. To identify this event, both a γ-ray from the annihilation of the positron and a γ ray from the nuclear capture of the neutron a sufficiently short time later must be detected. The imposition of both these requirements reduces sufficiently the background (determined by turning off the reactor) so that the rate of true antineutrino interactions is observable.

A convenient reaction through which to observe neutrinos is $\nu + Cl^{37} \rightarrow Ar^{37} + e^-$. If a reactor is used as a source of neutrinos (in fact, antineutrinos), this reaction is not observed, and the accuracy is sufficient thus to demonstrate that the antineutrino is distinct from the neutrino. Unfortunately the sensitivity of the experiment (R. Davis, 1955) is too small at present by a factor of 1000 to observe the expected neutrino flux from the sun, which emits neutrinos by the process



$$4p \rightarrow \text{(intermediate steps)} \rightarrow He^4 + 2e^+ + 2\nu$$

[C.J.C.]

Bibliography: J. S. Allen, *The Neutrino*, 1958

Neutron

An elementary particle having approximately the same mass as the proton, but lacking electric charge. It is indispensable in the structure of the elements, and in the free state it is an important reactant in nuclear research and the propagating agent of fission chain reactions.

Neutrons in nuclei. Neutrons and protons are the constituents of atomic nuclei. The role of neutrons in nuclei is in a way an indirect one, for although it is the number of protons in the nucleus that determines the chemical nature of an atom, nevertheless without neutrons, it would be impossible for two or more protons to exist stably together within nuclear dimensions, which are of the order of 10^{-12} cm. The protons, being positively charged, repel one another by virtue of their electrostatic interactions. The presence of neutrons weakens the electrostatic repulsion, without weakening the nuclear forces of cohesion. In light nuclei, the resulting balanced, stable configurations contain protons and neutrons in almost equal numbers, but in heavier elements, the neutrons outnumber the protons; in uranium-238 (U^{238}), for example, 146 neutrons are joined with 92 protons. Only one nucleus, hydrogen-1, contains no neutrons. For a given number of protons, neutrons in several different numbers within a restricted range often yield nuclear stability—hence the isotopes of an element.

Sources of free neutrons. Free neutrons have to be generated from nuclei, and since they are bound therein by cohesive forces, an amount of energy

Neuromeres are temporary constrictions in the hindbrain resulting from localized growth processes. They were once thought to have evolutionary significance as vestiges of metamerism, but the plasticity of the neural tube with respect to its mesodermal surroundings makes this interpretation doubtful.

Neural folds never form in teleost fish The neural plate concentrates into a solid neural keel which then hollows out secondarily as in the tail bud of other vertebrates. *See* EMBRYOLOGY; EMBRYONIC INDUCTION; GASTRULATION, NERVOUS SYSTEM; NEURAL CREST. [H.L.H.]

Neutralization

In chemical literature, processes in which the acidity or basicity of a solution is destroyed (neutralized) by the reaction of hydrogen ion with hydroxide ion to produce water.

When a solution containing exactly 1 mole of hydrochloric acid in aqueous solution is mixed with an aqueous solution containing exactly 1 mole of sodium hydroxide, the result is a solution containing 1 mole of the salt sodium chloride and the water which was present in both of the solutions before mixing plus 1 mole produced by the reaction. The resulting solution is said to be neutral because the hydrogen ion and hydroxide ion concentrations are equal

When, on the other hand, 1 mole of acetic acid is mixed with exactly 1 mole of sodium hydroxide, the final solution is slightly basic because the salt produced is hydrolyzed (*see* HYDROLYSIS). In this example neutralization does not produce a perfectly neutral solution. A common analytical procedure is the determination of the amount of an acid or base present by neutralization of some unknown amount of that substance with a carefully measured volume (or weight) of a standardized solution (one containing a known amount of reagent in a specified volume or weight). The end of such a titration is usually indicated by a very rapid change in the pH. A rapid change in pH can be detected accurately either by electrical methods or by indicators (chemical substances which change color when pH is altered within a fairly narrow range).

The word neutralize is sometimes used in the general sense of nullify. The effects of many chemicals or of change of conditions may be nullified by appropriate action of the experimenter, such as addition of other chemicals or appropriate change of conditions *See* ACID AND BASE; TITRATION. [T.F.Y.]

Bibliography: W. C. Pierce, E L Haenisch, and D. T. Sawyer, *Quantitative Analysis*, 1958.

Neutralization reaction (antibody)

A procedure in which the chemical or biological activity of a reagent or a living organism is inhibited, usually by a specific neutralizing antibody. *As an example, the lethal or the dermonecrotic actions* of diphtheria toxin on animals may be completely neutralized by an equivalent amount of diphtheria antitoxin—an antibody produced in animals or in humans after contact with diphtheria toxin or toxoid. Lesser amounts of antitoxin provide intermediate degrees of inhibition. These facts provide the basis for the Schick test for susceptibility *to diphtheria. Tetanus and botulinus toxins* may be similarly inhibited by their specific antitoxins. In contrast, the typical toxins of dysentery and other gram-negative bacteria are only slightly neutralized, even by large excesses of antibody. Antibodies to bacterial, snake venom, and other enzyme preparations regularly precipitate them from solution so that the supernates are devoid of enzyme activity; however, the neutralization of activity in the precipitate may range from complete to negligible. *See* BACILLARY DYSENTERY; BOTULISM; IMMUNOLOGY, NEUTRALIZING ANTIBODY; SEROLOGY, TETANUS.

Infection of a host by a living bacterium, virus, or other microorganism may also be inhibited or mitigated by the corresponding antibodies, and such neutralization tests are used in the diagnostic examination of sera or of infective agents recovered from such infections as poliomyelitis and yellow fever. In some instances, the antibody may be injected into a test animal before, or occasionally shortly after, challenge with the living agent. In other instances, the neutralization of the microbial infectivity by the antibody is permitted to take place in the test tube, and its degree determined by subsequent injection of the mixture into an appropriate test animal [H.P.T.]

Bibliography: W. C Boyd, *Fundamentals of Immunology*, 3d ed., 1956, T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.

Neutralizing antibody

An antibody that reduces or abolishes some biological activity of a soluble antigen or of a living microorganism. Thus, diphtheria antitoxin is a neutralizing antibody that, in adequate amounts, abolishes the pathological effects of diphtheria toxin in animals

Analogous neutralizing effects of antibodies can be demonstrated for the lytic effects of many lysins and, most important, for the pathogenic effects of viruses and the rickettsiae. Since the latter are complex bodies containing multiple antigens, not all of the resulting antibodies need have neutralizing activities, although they may display a variety of other serological properties Antibodies to enzymes constitute a special case; in all instances, they precipitate their corresponding enzyme, but the degree of neutralization may range from 0 to 100%. Antibodies to the endotoxins of the gram-negative bacteria regularly neutralize their toxicity only to a low degree. *See* ANTIGEN; ANTITOXIN; DIPHTHERIA; LYSIN; NEUTRALIZATION REACTION (*ANTIBODY*); *RICKETTSIOSES*; TOXIN, BACTERIAL. [H.P.T.]

Neutrino

A neutral particle having zero rest mass and spin ½(h/2π), where h is Planck's constant. It is emitted in β-decay and other weak decays. The existence of the neutrino was originally postulated to account for the energy, linear momentum, and angular momentum which is missing from the observed particles (the residual nucleus and the electron) resulting from a β-decay. The neutrino, denoted by the Greek letter ν, is a lepton which obeys Fermi-Dirac statistics, its only known interaction is the β interaction, which is so weak that the direct observation of the neutrino is very difficult. See ELEMENTARY PARTICLE; LEPTON; RADIOACTIVITY.

The antineutrino ($\bar{\nu}$) is distinct from the neutrino, it is an antilepton. A neutrino is emitted in positron β-decay (for example, $A^{37} \rightarrow Cl^{37} + e^{+} + \nu$), an antineutrino is emitted in electron β-decay (for example, $n \rightarrow p + e^{-} + \bar{\nu}$).

The neutrino is massless, like the photon, and travels at the speed of light. Its masslessness also implies that in a pure spin state its spin can be directed only along its motion or opposite to its motion. In the former case the neutrino is said to have a right-hand polarization, or briefly, to be right handed. (If the extended thumb of a right hand is taken to point in the direction of motion of the neutrino, the fingers, partially closed toward the palm, indicate the sense of spin of the right-handed neutrino.) Conversely, a neutrino whose spin is directed oppositely to its motion is polarized left handed. Because of its masslessness, the handedness of a neutrino is an intrinsic property, like its charge or rest mass. (The apparent handedness of a massive particle, however, depends on the observer. If the observer travels faster than the particle, he will see its mom[illegible]

[illegible]

[illegible] ...und by experiment that only left-handed neutrinos exist, and that the antineutrino is right handed. Thus the neutrino field has only two possible quanta (left-handed, ν and right handed, $\bar{\nu}$), rather than the four ordinarily possible for a spin ½ particle (right- and left handed particle and antiparticle). Therefore the neutrino field need only have two components, rather than the four specified by the Dirac equation (see QUANTUM FIELD THEORY). This [illegible]

[illegible]

[illegible] violated. For upon inversion of space, the neutrino becomes right handed, which thus distinguishes the inverted world. T. D. Lee and C. N. Yang exploited this fact in their theory of parity nonconservation in weak interactions, conjecturing that only neutrinos of one handedness actually exist. See PARITY (QUANTUM MECHANICS); SYMMETRY LAWS (PHYSICS).

Inverse beta decay. This provides the direct evidence for the neutrino and was observed by F. Reines and C. L. Cowan in spite of the exceedingly small interaction cross section of the neutrino, ~10^{-43} cm². A nuclear reactor is a strong source of antineutrinos, which may undergo the reaction $\bar{\nu} + p \rightarrow n + e^{+}$. To identify this event, both a γ ray from the annihilation of the positron and a γ-ray from the nuclear capture of the neutron a sufficiently short time later must be detected. The imposition of both these requirements reduces sufficiently the background (determined by turning off the reactor) so that the rate of true antineutrino interactions is observable.

A convenient reaction through which to observe neutrinos is $\nu + Cl^{37} \rightarrow Ar^{37} + e^{-}$. If a reactor is used as a source of neutrinos (in fact, antineutrinos), this reaction is not observed, and the accuracy is sufficient thus to demonstrate that the antineutrino is distinct from the neutrino. Unfortunately the sensitivity of the experiment (R. Davis, 1955) is too small at present by a factor of 1000 to observe the expected neutrino flux from the sun, which emits neutrinos by the process



$$4p \rightarrow (\text{intermediate steps}) \rightarrow He^{4} + 2e^{+} + 2\nu$$

[C.J.C.]

Bibliography. J. S. Allen, *The Neutrino*, 1958.

Neutron

An elementary particle having approximately the same mass as the proton, but lacking electric charge. It is indispensable in the structure of the elements, and in the free state it is an important reactant in nuclear research and the propagating agent of fission chain reactions.

Neutrons in nuclei. Neutrons and protons are the constituents of atomic nuclei. The role of neutrons in nuclei is in a way an indirect one, for although it is the number of protons in the nucleus that determines the chemical nature of an atom, nevertheless without neutrons, it would be impossible for two or more protons to exist stably together within nuclear dimensions, which are of the order of 10^{-13} cm. The protons, being positively charged, repel one another by virtue of their electrostatic interactions. The presence of neutrons weakens the electrostatic repulsion, without weakening the nuclear forces of cohesion. In light nuclei, the resulting balanced, stable configurations contain protons and neutrons in almost equal numbers, but in heavier elements, the neutrons outnumber the protons; in uranium-238 (U^{238}), for example, 146 neutrons are joined with 92 protons. Only one nucleus, hydrogen-1, contains no neutrons. For a given number of protons, neutrons in several different numbers within a restricted range often yield nuclear stability—hence the isotopes of an element.

Sources of free neutrons. Free neutrons have to be generated from nuclei, and since they are bound therein by cohesive forces, an amount of energy

equal to the binding energy must be expended to get them out. Conversely, upon capture by a nucleus, the binding energy is released, and appears as capture γ-rays. Usually the binding energy for each neutron amounts to 6–8 Mev (*see* BINDING ENERGY, NUCLEAR; NUCLEAR REACTION). Nuclear machines such as cyclotrons and electrostatic generators induce many nuclear reactions when their ion beams strike target material; among the reactions are almost inevitably some which release neutrons, and these machines are sources of high neutron flux.

A portable type of neutron source consists of radium mixed with beryllium. The radium emits alpha particles, and when these strike beryllium nuclei, neutrons are released in the reaction $Be^9(\alpha,n)C^{12}$. This is the reaction with which the neutron was discovered in 1932. A source containing 1 g of radium bromide in a pellet with beryllium powder, suitably encased, measures about 1 in. in diameter by 1 in. long, and emits about 10^7 neutrons per second. Another compact source takes advantage of the fact that the γ-rays emitted by a number of radioactive nuclei exceed in energy the exceptionally low binding energy of neutrons in beryllium (1.67 Mev) and deuterium (2.23 Mev). For example, the illustration shows a piece of radioactive antimony encased in beryllium. Some of the 1.70-Mev γ-rays resulting from the decay of antimony interact with beryllium nuclei to release neutrons, which are generated with an energy of $1.70 - 1.67 = 0.03$ Mev. Neutrons liberated by the action of γ-rays or x-rays are called photoneutrons.

Neutrons are released in the act of fission, and nuclear reactors are unexcelled as intense neutron sources. The absorption of one neutron by a uranium-235 (U^{235}) nucleus is required to induce fission, but 2.5 neutrons are on the average released; this regeneration makes possible the nuclear chain reaction. A powerful research reactor may generate neutrons in such abundance that 1 cm^2 of a sample placed therein would be traversed by 10^{14} neutrons per second. A hole through the surrounding shield can yield a collimated beam having a unidirectional flux of 10^8 neutrons/(cm^2)(sec). The explosion of a 10-kiloton nuclear bomb releases about 10^{30} neutrons in about 1 μsec. *See* ATOMIC BOMB; CHAIN REACTION, NUCLEAR; FISSION, NUCLEAR.

Neutrons occur in cosmic rays, being liberated from atomic nuclei in the atmosphere by collisions of the high-energy primary or secondary charged particles. They do not themselves come from outer space (*see* COSMIC RAYS). For information on neutrons of another origin, *see* NEUTRON, DELAYED.

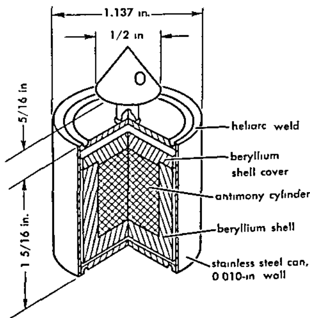


Diagram of antimony-beryllium photoneutron source of the type distributed by the Atomic Energy Commission from the Oak Ridge National Laboratory. The cone at the top is for remote handling.

Penetrating power. Neutrons resulting from nuclear reactions usually possess kinetic energies of the order of 1 Mev. Having no electric charge, they interact so slightly with atomic electrons in matter that energy loss by ionization and atomic excitation is essentially absent. Consequently, they are vastly more penetrating than charged particles of the same energy. The main energy loss mechanism occurs when they strike nuclei. As with rolling balls, the most efficient slowing-down occurs when the bodies that are struck in an elastic collision have the same mass as the moving bodies; hence, the most efficient neutron moderator is hydrogen, followed by other light elements: deuterium, beryllium, and carbon. The struck nucleus loses energy by ionizing surrounding atoms, eventually producing heat, and in living tissue, biological damage.

The great penetrating power of neutrons imposes severe shielding problems for reactors and other nuclear machines, and it is necessary to provide walls, usually of concrete, several feet in thickness to protect personnel. The currently accepted health tolerance levels for a 40-hour week are, for fast neutrons, 40 neutrons/(cm^2)(sec); for slow neutrons, 2000/(cm^2)(sec). *See* RADIATION INJURY (BIOLOGY); RADIATION SHIELDING.

Detection of neutrons. In pulse counting, neutrons are allowed to produce exothermic (energy-releasing) nuclear reactions, the ionizing products of which are made to generate electrical impulses that can be amplified for individual counting. A proportional counter containing boron, either as a coating on the inner walls or as a filling gas (boron trifluoride), counts neutrons by virtue of the reaction $B^{10}(n,\alpha)Li^7 + 2.78$ Mev. An ionization chamber coated internally with U^{235} gives ionization pulses from the energy of fission fragments as they travel through the gas. A lithium iodide crystal (europium activated) scintillates because of the energy released by the reaction $Li^6(n,\alpha)H^3 + 4.78$ Mev. The light pulses (scintillations) are reflected onto a photomultiplier, which transforms them to electrical pulses. Capture γ-rays emitted from strong neutron absorbers such as cadmium can similarly be registered by scintillation countin-

Large and sensitive neutron detectors have been made by dissolving cadmium or boron salts in tanks containing scintillating liquids *See* SCINTILLATION DETECTOR, LIQUID, *see also* PARTICLE DETECTOR.

In detection by activation, advantage is taken of the fact that many elements become radioactive under neutron irradiation A sample is exposed, [illegible]

If the neutrons are originally fast, the foregoing methods gain sensitivity if the detector is surrounded by a few inches of hydrogenous material, such as paraffin, for this moderates (decreases) the speed of the neutrons and makes their capture by nuclei more probable.

Detection by recoil is particularly applicable to the counting of fast neutrons A counter with hydrogenous walls or filling gas, for example, methane, gives pulses because the protons produce ionization when they recoil after being struck by the fast neutrons

Intrinsic properties. Free neutrons are themselves radioactive, each transforming spontaneously into a proton, an electron (β^- particle), and an antineutrino The energy release is 0.786 Mev per event, and the half-life is 11.3 ± 0.5 min This instability is a reflection of the fact that neutrons are slightly heavier than hydrogen atoms The neutron's rest mass is 1.008982 atomic mass units, (1.67474×10^{-24} g), as compared with 1.008142 [illegible]

[illegible] ...neutrons, that is, beams of neutrons the magnetic dipoles of which are aligned predominantly parallel to one direction in space. The magnetic moment is −1.913141 nuclear magnetons. The magnetic structure has a finite size, being roughly exponential in intensity with a root-mean square radius of 0.9×10^{-13} cm Neutrons spin with an angular momentum of ½ in units of $h/2\pi$, where h is Planck's constant The negative sign attached to the magnetic moment indicates that the magnetic moment vector and the angular momentum vector are oppositely directed *See* ANTINEUTRON; ELEMENTARY PARTICLE; NEUTRON CROSS SECTION; NEUTRON DIFFRACTION; NUCLEAR STRUCTURE; THERMAL NEUTRONS. [A H S]

Bibliography: L. F. Curtiss, *Introduction to Neutron Physics*, 1959.

Neutron, delayed

A neutron emitted spontaneously from a nucleus as a consequence of excitation left from a preceding radioactive-decay event Delayed neutrons are of interest as an unusual phenomenon in radioactivity and are of practical importance in the control of nuclear chain reactors

Radioactive transformation by β-decay often leaves a product nucleus with internal energy in excess of that associated with its stable state, and customarily the energy is radiated as γ-ray quanta. In exceptional cases the energy may exceed that required to remove one of the constituent neutrons from the product nucleus; when this happens, spontaneous neutron emission takes place and serves very rapidly to de-excite the nucleus. The delayed neutrons are accordingly a kind of radioactivity, observable at the time of decay of the β-emitting precursor, and since the appearance of each neutron is delayed by the slowness of the preceding β-decay, the half lives and the chemical behavior of the delayed neutron radioactivities are in practice those of the precursors.

In fission, the delayed neutrons afford a contrast with the prompt-fission neutrons, the latter cease when fissioning ceases, but delayed neutrons are emitted from fission products for minutes thereafter. The energetic conditions required for delayed neutron emission are most likely to be fulfilled when the neutron is weakly bound to its containing nucleus, and hence the chemical occurrence of delayed neutron emission is related to the neutron shell structure of nuclei.

The table enumerates the delayed neutron emitters now known. For the importance of delayed

Characteristics of delayed neutron emitters

Precursor (β^--emitter)	Half-life of precursor, sec*	Product of β^- emission (delayed neutron emitter)	Product of delayed neutron emission	Remarks
Li^9	0.168	Be^9	$Be^8 \rightarrow 2He^4$	Not a fission product
N^{17}	4.14	O^{17}	O^{16}	Not a fission product
Br^{87}	54.2	Kr^{87}	Kr^{86}	Fission product
Br^{88}	16.3	Kr^{88}	Kr^{87}	Fission product
Br^{89}	4.4	Kr^{89}	Kr^{88}	Fission products; mass assignment surmised
Br^{90}	1.6	Kr^{90}	Kr^{89}	
I^{137}	24.4	Xe^{137}	Xe^{136}	Fission product
I^{138}	6.3	Xe^{138}	Xe^{137}	Fission product
I^{139}	2.0	Xe^{139}	Xe^{138}	Fission product
?	0.6	?	?	Fission products, may be a complex of activities
?	0.23	?	?	

* Half-lives of precursors are effective half-lives of delayed neutron radioactivities

neutrons in the control of fission reactors, *see* REACTOR PHYSICS. *See also* FISSION, NUCLEAR; NEUTRON; RADIOACTIVITY. [A.H.S.]

Bibliography: D. J. Hughes et al. (eds.), Progress in Nuclear Energy, Series 1: *Physics and Mathematics*, vol. 1, 1956; A. F. Stehney and G. J. Perlow, *Proc. 1958 Geneva Conference on Peaceful Uses of Atomic Energy*, Paper 15(P)691, 1959.

Neutron cross section

The effective target area presented by a nucleus to *an incident neutron, expressing the probability* that an interaction of a given kind will take place. Neutron cross sections are commonly expressed in units of 10^{-24} cm², or barns. They enter into all quantitative considerations of neutron-nucleus interactions.

Description. *Consider a parallel beam of neutrons* impinging perpendicularly upon matter, as in Fig. 1. Let the beam have a cross-sectional area of 1 cm²; let the density of neutrons in the beam be n per cm³, and let their velocity be v. Then a flux nv of neutrons will fall upon the sample per square centimeter of surface area per second. Many of the neutrons will pass straight through the sample, but some will not; the beam will be attenuated by the processes of nuclear absorption and scattering. Considering the neutrons as infinitesimal, suppose that each nucleus in the sample presents a target area, σ, without overlap, such that a collision will remove that neutron from the beam, and suppose that the sample is thin enough to make second collisions improbable. Then, if there are N nuclei per cm³ in the sample, and if the thickness of the sample is Δx cm, the opacity of the sample will be expressed by the product $\sigma N \Delta x$. The number of neutrons lost per second from the beam will be proportional to this and to the incident flux:

$$-\Delta(nv) = \sigma N \Delta x \cdot (nv) \qquad (1)$$

whence, by integration, the attenuation factor is seen to be $e^{-N\sigma\Delta x}$. The quantity σ is called the cross section; *in this case it is the total cross section inasmuch* as it expresses the effective target area per nucleus for removal of neutrons from the beam without specifying the process involved in the removal.

Generalization of concept. Another way to look at Eq. (1) is to consider that σ is a constant of proportionality, with dimensions of an area, which relates the number of removal events $\Delta(nv)$ with the number of target nuclei $N\Delta x$ and the neutron flux nv. This broader concept enables one to generalize and to apply the relationship to a situation in which neutrons are going in all directions, provided that nv is then understood to designate the number of neutrons per second traversing a sphere with a maximum cross-sectional area of 1 cm². If more than one process can function in removing neutrons from the beam, the total cross section can be subdivided; for example, if absorption and elastic scattering are the only two processes in question, then $\sigma_{total} = \sigma_{abs} + \sigma_{scat}$. If scattering is

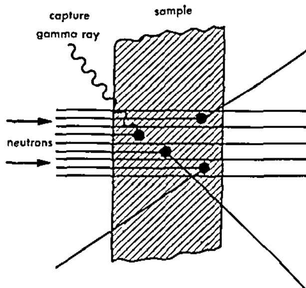


Fig. 1. The concept of effective target area in collisions between neutrons and atomic nuclei in matter. Three neutrons are scattered, and one is absorbed by the sample, the transmitted neutron beam is thus attenuated.

three times as probable as absorption, then $\sigma_{scat} = 3\sigma_{abs}$. Cross-section values can be attached to all forms of interaction of neutrons with nuclei: elastic scattering, absorption, inelastic scattering, special reactions such as fission, (n,p), $(n,2n)$, etc. *In the case of scattering, cross sections* can be used to express the probability of deflection through a given angle; such cross sections are called differential cross sections. Integration of the differential cross section over the solid angle yields the cross section for scattering.

The concept of cross sections is not necessarily related to the geometrical size of the nucleus. This would indeed be the case if the nuclei were hard spheres, in which case the effective target area would be πr^2, with r designating the radius of the nucleus. Actually, the wave character of both the nucleus and the neutron modifies the interaction so strongly that the various cross sections often vary sharply with neutron energy, going through maxima and minima in a sometimes spectacular manner. For neutrons of a given energy—for example, thermal neutrons (neutrons in thermal equilibrium with the substance in which they exist)—the ab*sorption cross sections vary over a tremendous* range in magnitude, as shown in the table, and change unpredictably from element to element.

The practical usefulness of cross-section information lies in the reverse use of Eq. (1), for, knowing the cross section for a given process, one can calcu-

Thermal neutron absorption cross sections for a neutron velocity of 2200 m/sec (energy: 0.023 ev)

Nuclide	σ_{abs}, barns	Nuclide	σ_{abs}, barns
H^1	0.332	Cd^{113}	20,000
Be^9	0.010	I^{127}	7.0
B^{10}	3813	Xe^{135}	2,720,000
N^{15}	0.000024	Bi^{209}	0.034

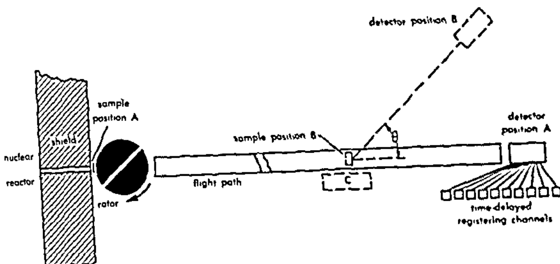


Fig 2 Schematic diagram of a neutron chopper and time-of flight apparatus.

late directly the number of events that will be produced in a given sample when it is placed in a given neutron flux The scientific usefulness lies in the fact that cross sections are a meeting ground for theory and experiment

Measurement. One important method of measuring cross sections involves the use of a pulsed neutron source together with time-of flight techniques, in which the difference in velocities of neutrons in known, narrow energy bands is used to determine the energies In the example to be discussed, a neutron chopper is used to produce the pulses of neutrons.

At the left (Fig. 2), a narrow slit through the shielding wall of a reactor permits a collimated beam of neutrons of mixed energy to emerge, flying toward the right Just outside the shield, a heavy metal rotor with one or more diametrical slots (Fig 3) is spun so that when the slot is in line with the collimator, a pulse of neutrons can pass through The neutrons then traverse a flight path, perhaps 100 m long and the faster neutrons in the pulse reach the detector at A before the slower neutrons The detector, typically a boron-containing counter or scintillator, generates electrical pulses promptly when neutrons fall upon it The pulses are amplified and routed to a series of registers, which are arranged to record at successive time intervals following the instant of opening of the rotor, for example, channel 13 might register neutrons that take from 5.0 to 5.2 μsec to travel from rotor to detector, and so on As the rotor spins, the counts in the various channels accumulate, and the registers record a representation of the velocity spectrum of the neutrons coming from the reactor For measurement of total cross section, a sample is introduced at A, it causes attenuation of the transmitted neutron beam by the factor $e^{-N\sigma\Delta x}$ If σ is particularly large for any one neutron velocity then the time channels associated with that velocity will receive fewer counts, and a dip will appear in the registered velocity spectrum The dips are interpreted inversely as peaks in σ; $N\Delta x$ is determined by weighing the sample, and thus measurement of $e^{-N\sigma\Delta x}$ suffices to determine σ absolutely. The data of Fig 4 were obtained in this way

The time-of flight method is capable in principle of measuring differential, scattering, and absorption cross sections also If the sample and a second detector are at positions B of Fig. 2, then the detector will register neutrons scattered through the angle θ If the detector is provided with time-delayed counting channels like detector A, and if the relative efficiencies of detectors A and B are known, then one derives σ_{dif} from the fraction of neutrons of a given velocity that are scattered by the known number of atoms in the sample. The value of σ_{scat} can be obtained by integration of σ_{dif} over the scattering angle, or alternatively, detector B can be made so that it nearly surrounds the sample To obtain σ_{abs}, a scintillator at position C in Fig 2, with time-delayed channels, can register the capture γ-rays produced when neu-

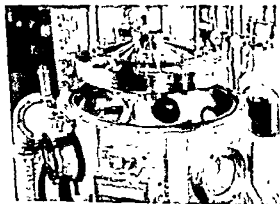


Fig 3 *Photograph* of chopper rotor at Oak Ridge National Laboratory. The rotor is shown suspended over its housing, which will later be *covered* and evacuated to reduce air drag The axis of spin is vertical. Diametrical slots *in this* rotor are arranged in groups, one of which can be seen at the center.

trons are absorbed. If the geometrical interception factor and the efficiency of the gamma counter are known, the fraction of neutrons of any velocity that are absorbed by the known number of atoms in the sample can be derived, and hence σ_{abs} may be obtained.

The time-of-flight method can use cyclotrons, linear accelerators, or high-voltage generators as pulsed neutron sources instead of a chopper. In various embodiments, this method can cover the neutron energy range from roughly 10^{-3} ev to 10^{6} ev or more. It is, however, only one of a number of neutron-energy selection methods that are used in cross-section investigations.

Energy dependence. At neutron energies of less than about 1 *ev*, σ_{total} for many nuclides shows a regular behavior, frequently varying inversely as *the velocity of the neutron For boron-10, this 1/v* dependence is followed from 0.001 to 100 ev. At energies in the range 10 to 10^{6} or 10^{7} ev, the cross sections for nearly all nuclides exhibit remarkable *peaks and valleys, known as resonances.* Figure 4 shows resonances for neptunium-237. There are resonances both in σ_{abs} and in σ_{scat}, which may or may not occur at the same energy. At neutron energies of the order of 10^{9} ev, the resonances fade out.

The heights, widths, and spacings of resonances are revealing in nuclear theory, because resonances are associated with nuclear energy levels. Absorption of a zero-energy neutron into a nucleus with atomic number Z and weight A results in the nucleus Z^{A+1} in a state having surplus energy corresponding to the binding energy (*see* BINDING ENERGY, NUCLEAR) Suppose, for example, that this surplus energy amounts to exactly 6,000,000.0 ev. *At this state of excitation,* nuclear energy levels are abundant and narrowly spaced. If there happens to be a level at 6,000,000.5 ev, then it would be matched exactly if the incident neutron carried with it 0.5 ev of kinetic energy. An interaction *would be especially probable, as in the first peak in* Fig. 4, and one would say that nucleus (Z^{A}) shows a resonance at a neutron energy of 0 5 ev. The resonances, which can be examined in great detail by the methods of cross-section measurements, are accordingly directly informative about the energy-level structure of nuclei. *See* NEUTRON; REACTOR PHYSICS; SCATTERING EXPERIMENTS, NUCLEAR.

[A H.S.]

Bibliography: D. J. Hughes, *Pile Neutron Research*, 1953; D. J. Hughes and R. B. Schwartz, *Neutron Cross Sections*, 2d ed., 1958; J. B. Marion and J. L. Fowler (eds.), *Fast Neutron Physics*, 1959.

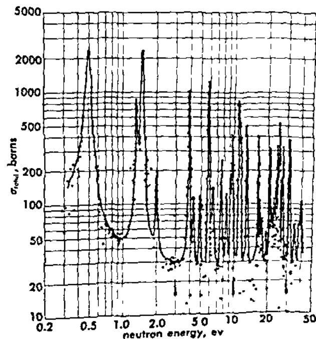


Fig 4. Variation of σ_{total} with neutron energy for neptunium-237, showing resonances The data were taken with the apparatus shown in Fig 3.

Neutron diffraction

The phenomenon associated with the *interference* processes which occur when neutrons are scattered by the atoms within solids, liquids, and gases The use of neutron diffraction as an experimental technique is relatively new compared to electron and *x-ray diffraction, since successful* application requires high thermal neutron fluxes which can be obtained only from nuclear reactors (A thermal neutron is defined as a neutron possessing a kinetic energy of about 0.025 electron volts) Although experiments have been performed which contributed basic information in nuclear physics, the most important contributions from this technique are in solid-state studies of magnetism and the structure of matter These investigations are possible because the thermal neutrons from nuclear reactors have *energies with equivalent wavelengths near* 1 A and are therefore ideally suited for interatomic interference studies. In its applications to solid-state problems, neutron diffraction is very similar in both theory and experiment to x-ray diffraction, but its importance arises from the significant differences in the scattering of these two types of radiation.

The scattering of x-rays by atoms results from a scattering interaction with the atomic electrons, and the scattering amplitudes are approximately [illegible] the atomic number of the scatterer [illegible]

as a form factor, that is descriptive of the spatial distribution of the electrons *In the scattering* of neutrons by atoms, there are two important interactions. One is the short range, nuclear interaction of the neutron with the atomic nucleus This interaction produces isotropic scattering, because the nucleus is essentially a point scatterer relative to the wavelengths of thermal neutrons There is no regular variation of the nuclear scattering amplitudes with atomic number because strong reso-

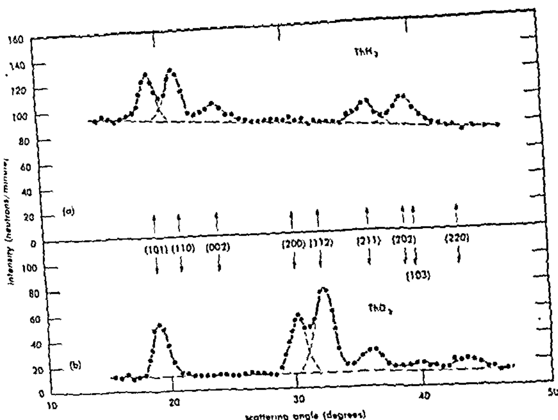


Fig. 1. Neutron diffraction patterns from (a) polycrystalline thorium hydride, ThH_2, and (b) polycrystalline thorium deuteride, ThD_2. Differences in the patterns are caused primarily by differences in the nuclear scattering from hydrogen and deuterium atoms.

nances are associated with the scattering process. Moreover these resonances affect the phase changes between the incident and scattered neutron waves, so that the scattering amplitudes can be positive or negative. The second process for the scattering of neutrons by atoms is the interaction of the magnetic moment of the neutron with the spin and orbital magnetic moments of the atom, and therefore this magnetic scattering occurs only when the scattering specimen possesses an atomic magnetic moment. The amplitude of the interaction varies with the size and orientation of the atomic magnetic moment, and the intensity of scattering has a form factor angular dependence that is representative of the magnetic electrons.

Techniques. Although thermal neutron beams from nuclear reactors have intensities that are lower than those obtained from efficient x-ray tubes, most of the x-ray diffraction methods can be used with neutrons. Furthermore, since the neutron absorption cross section for many materials is very small, diffraction effects can also be investigated by observations of the neutrons transmitted through a sample. In most experiments, however, the sample is irradiated with monochromatic neutrons, and the scattered radiation is measured with a neutron detector such as a proportional counter filled with boron trifluoride, BF_3, gas. In structure determinations, only the angular distribution of the scattered neutrons is required; but in inelastic scattering experiments, that is, experiments involving a change in the neutron energy, the energy distribution must also be measured with an additional crystal spectrometer or a neutron velocity selector. Both polycrystalline and single crystal specimens can be examined, and auxiliary equipment for controlling the sample conditions can be constructed easily because of the relatively low neutron cross sections. Since the thermal neutrons from a reactor have a continuous energy distribution with no pronounced peaks, the monochromatic beams used in these experiments must be obtained by isolating a narrow slice of the neutron spectrum. This is usually accomplished by diffraction of the reactor neutrons from large single crystals, but filters and neutron velocity selectors can also be used. In certain investigations of magnetic scattering, polarized neutron beams are required. Such beams can be obtained in the monochromating process involving diffraction from single crystals, because the neutrons scattered under specific conditions from particular ferromagnetic crystals are almost completely polarized.

Chemical crystallography. Since the nuclear scattering amplitudes for neutrons do not vary uniformly with atomic number, there are certain types of chemical structures which can be investigated more readily by neutron diffraction than by x-ray diffraction. Moreover, since neutron scattering is a nuclear process, a particular isotope can frequently

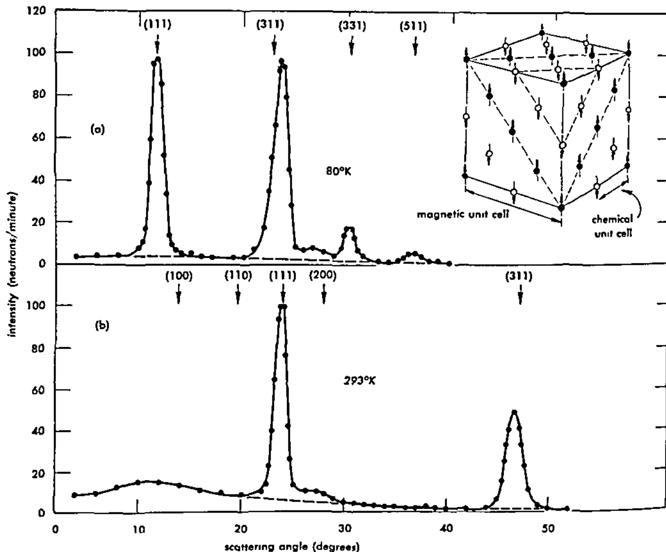


Fig. 2. Neutron diffraction patterns from polycrystalline manganese oxide, MnO, at temperatures (a) below and (b) above the antiferromagnetic transition at 122°K. At 293°K, only nuclear reflections are observed, while at 80°K, additional reflections are obtained from the indicated antiferromagnetic structure. The atomic magnetic moments in this structure are directed along a magnetic axis within the (111) planes.

be substituted for the normal element if it has a more favorable scattering amplitude. The most important application of neutron diffraction in chemical crystallography is the structure determination of composite crystals which contain both heavy and light atoms, such as compounds containing hydrogen. Since hydrogen and deuterium have neutron scattering amplitudes which are comparable to those of other atoms, their positions in crystals can be determined by this technique, whereas x-ray diffraction usually gives little information about them (see Fig. 1). The order and disorder in many alloy systems comprising atoms with almost the same atomic number can also be determined. Furthermore, since the scattering of neutrons by the nucleus is isotropic, the neutron technique is advantageous in investigations of liquids, gases, amorphous materials, and other structures where the features of the diffraction pattern at large scattering angles are significant.

Magnetic scattering. The magnetic *scattering of* neutrons furnishes information on the magnetic properties of the individual atoms in a material and offers a unique approach to the study of magnetic phenomena. Each type of magnetic lattice in a solid which displays magnetic properties has a characteristic diffraction pattern from which the magnitude and specific orientation of the atomic magnetic moments can be determined. For paramagnetic materials, where the atomic moments are uncoupled and randomly oriented in direction, the magnetic scattering is diffuse. For ordered magnetic lattices, the magnetic scattering is found in Bragg reflections. (For a discussion of Bragg reflections, see X-RAY DIFFRACTION.) Magnetic reflections from ferromagnetic materials occur superimposed on the nuclear reflections, but for antiferromagnetic materials, in which the atomic moments are oriented with no net magnetization per unit volume, superlattice reflections are observed at other angles, as shown in Fig. 2. Since ferrimagnetic materials have atomic moments with antiparallel components but also possess a net ferromagnetic moment, magnetic reflections are observed at both nuclear and other positions. Consequently, neutron diffraction investigations at various sample

temperatures can determine the ordering transition temperature, the type of magnetic ordering, and the nature of the magnetic coupling which exists in the ordered lattice. Furthermore, the form factor for the magnetic scattering of neutrons can be interpreted in terms of the spatial distribution and angular momentum characteristics of the magnetic electrons within the atoms. Information can also be obtained on ferromagnetic and antiferromagnetic domains and on the magnetic anisotropy which exists within magnetic structures. *See* ANTIFERROMAGNETISM, FERRIMAGNETISM, FERROMAGNETISM, PARAMAGNETISM

Inelastic scattering. Diffraction effects caused by the inelastic scattering of neutrons are more pronounced than those observed in x-ray scattering because of the different momentum-energy ratios. Moreover, with thermal neutrons, the energy changes can occur either by an interaction with the lattice vibrations or by a magnetic interaction with the atomic magnetic moments. Analyses of inelastic neutron scattering by the former process can be interpreted directly in terms of the dispersion relations of the normal modes of the crystal and do not require the large corrections necessary in similar x-ray investigations. Analyses of the inelastic interaction of neutrons with magnetic spin-waves in crystals are capable of giving important information on the energy levels which exist in magnetic materials. *See* ELECTRON DIFFRACTION, NEUTRON CROSS SECTION [M.K.W.]

Bibliography. G. E. Bacon, *Neutron Diffraction*, 1955; D. J. Hughes, *Neutron Optics*, 1954; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 2, 1956

Neutron optics

A title-by-analogy of certain phases of neutron physics in which the wave character of neutrons dominates and leads to behavior similar to that of light. Neutrons can be reflected at small glancing angles from plane surfaces, they show various scattering phenomena with similarity to light, and they can be diffracted by crystals (*see* NEUTRON DIFFRACTION). Although they can also be polarized, the analogy with light is in this case invalid because the polarization of neutrons depends upon their possession of a constant magnetic moment, which light waves lack. *See* NEUTRON. [A.H.S.]

Bibliography. D. J. Hughes, *Neutron Optics*, 1954

Newcastle disease

An epizootic viral disease of fowls, with respiratory, gastrointestinal, and central nervous system involvement; it may be transmitted to human beings who work with fowls, usually appearing as a conjunctivitis.

The virus is related to influenza virus and other myxoviruses in size, host range, and hemagglutination characteristics. *See* INFLUENZA, MYXOVIRUS.

The disease in adult fowls is influenzalike; in young ... moencephalitis is frequent. Mortality rates vary. A live, attenuated virus is available for prevention of the disease in birds. *See* BIOLOGICALS. [J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.

Newton

A unit of force in the meter-kilogram-second system of units. One newton is the *force which will impart* 1 m/sec² acceleration to the International Prototype Kilogram mass. The International Prototype Meter is the standard unit of length. *See* FORCE. [C.T.P.]

Newton's laws of motion

Three fundamental principles which form the basis of classical, or Newtonian, mechanics. They are stated as follows:

First law. A particle not subjected to external forces remains at rest or moves with constant speed in a straight line.

Second law. The acceleration of a particle is directly proportional to the resultant external force acting on the particle and is inversely proportional to the mass of the particle.

Third law. If two particles interact, the force exerted by the first particle on the second particle (called the action force) is equal in magnitude and opposite in direction to the force exerted by the second particle on the first particle (called the reaction force).

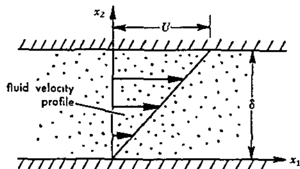
The first law, sometimes called Galileo's law of inertia, can now be regarded as contained in the second. At the time of its enunciation, however, it was important as a negation of the Aristotelian doctrines of natural placement and continuing force.

The third law, sometimes called the law of action and reaction, was also to some extent established prior to Newton's statement of it. However, Newton's formulation of the three laws as a mutually consistent set, with the nature of force clearly defined in the second law, provided the basis for classical dynamics.

The Newtonian laws have proved valid for all mechanical problems not involving speeds comparable with the speed of light (approximately 300,000 kilometers/sec) and not involving atomic or subatomic particles. The more general classical methods of Lagrange and Hamilton are elaborations of Newtonian principles. *See* HAMILTON'S EQUATIONS OF MOTION; LAGRANGE'S EQUATIONS; *see also* DYNAMICS; FORCE, KINETICS (CLASSICAL MECHANICS); MOTION. [D.W.I.]

Newtonian fluid

A fluid in which the state of stress at any point is a linear function of the time rate of strain at that point. The fluid thus bears a direct analogy to a Hookean solid, for which the state of stress is a linear function of the strain. Many gases and liquids are closely Newtonian over a wide range of pressures and temperatures.



Top plate moves relative to bottom plate to produce Couette flow of intervening viscous fluid

The simplest example of Newtonian fluid flow is Couette flow, the low-speed steady motion of a viscous fluid between two infinite plates moving parallel to each other with relative velocity U as illustrated The shear stress τ_{21} in the fluid is constant and equals $\mu(\partial u_1/\partial x_2) = \mu(U/\delta)$ The time rate of strain at a point is a tensor quantity

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

where u_i are velocity components of fluid and x_i are rectangular Cartesian coordinates with $i = 1,2,3$ Fluids are inherently isotropic so that the most general linear relationship between stress τ_{ij} and ε_{ij} is

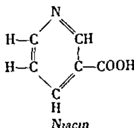
$$\tau_{ij} = -P\,\delta_{ij} + (\lambda - \tfrac{2}{3}\mu)\,\delta_{ij}\varepsilon_{mm} + 2\mu\varepsilon_{ij}$$

where μ is the ordinary viscosity coefficient, λ is the bulk or volume viscosity coefficient, and P is the pressure In gases, $\lambda = 0$ if the molecules have no internal degrees of freedom or if the internal motions are not excited. For low Mach numbers, $\varepsilon_{mm} = 0$, so λ does not appear in the stress relationship. For most liquids, μ decreases with temperature and increases with pressure; for gases it increases with temperature and is almost independent of pressure. *See* FLUID-FLOW PRINCIPLES

[A.E.BR.]

Niacin

A vitamin also known as nicotinic acid. It is a white water-soluble powder stable to heat, acid, and alkali, with the following structural formula



Niacin

It is found in biochemically active combinations

cells studied have enzymic systems involving niacin.

Many animals, including man, are capable of synthesizing niacin in varying degrees from the amino acid tryptophan. Niacin is widely distributed in foods. Yeasts, wheat germ, and meats, particularly organ meats, are rich sources of the vitamin. Some foods such as milk are relatively poor sources of niacin but contain generous quantities of tryptophan.

Niacin-deficiency disease is known as pellagra and is particularly prevalent among the poor people whose diet is largely corn. Pellagra is characterized by dermatitis, dementia, diarrhea, and death. Skin lesions are usually observed in areas exposed to the sun. The disease is accompanied by large gastrointestinal lesions

Niacin is present in enzymes in the form of two coenzymes, diphosphopyridine nucleotide (DPN) or coenzyme I and triphosphopyridine nucleotide (TPN) or coenzyme II. Enzymes containing DPN or TPN function in oxidation-reduction systems by virtue of their ability to accept hydrogen ions (protons) and electrons from substrates and transfer them to other hydrogen acceptors, such as the flavoproteins (*see* RIBOFLAVIN). Niacin-containing enzymes catalyze about 40 reversible biochemical reactions, many of different types, as illustrated by the following: acetaldehyde ⇌ ethanol; 1,3-diphosphoglyceric acid ⇌ 3-phosphoglyceraldehyde; pyruvic acid ⇌ lactic acid; imino glutaric acid ⇌ L-glutamic acid; acetic acid ⇌ acetaldehyde.

Unlike the other B vitamins, little niacin is excreted in the urine Excess niacin is excreted in man mostly as N^1-methylnicotinamide and the 6-pyridone of N^1-methylnicotinamide. The pellagra-preventive potency of a diet is related not only to its niacin and tryptophan content, but also to the availability of its niacin. There is evidence that some of the niacin of foods cannot be released by digestive enzymes. The existence of an antiniacin material in corn has been suggested. The effect of the carbohydrate content of the diet on the synthesis of niacin by intestinal bacteria may also be of some importance The use of urinary excretion data to determine nutritional status with regard to niacin has been disappointing. The recommended dietary allowances of the National Research Council for niacin are 10 times the thiamine allowances. This provides considerably more than average needs *See* DIPHOSPHOPYRIDINE NUCLEOTIDE (DPN); TRIPHOSPHOPYRIDINE NUCLEOTIDE (TPN); VITAMIN [S N G.]

Industrial production. Niacin is produced to U.S. Pharmacopeia (USP) requirements The principal route for manufacture of niacin is oxidation of quinoline, obtained from coal tar (Fig 1). Oxidation of 2-methyl-5 ethylpyridine is nearly as important; other processes involve oxidation of 3-picoline, hydrolysis of 3-cyanopyridine, or oxidation and hydrolysis of nicotine. Many patents cover the field. Nicotinamide is produced by amidation of niacin or its esters and by hydrolysis of 3-cyanopyridine.

Over 3,000,000 lb of niacin were produced in 1957 in the United States; less ... rd

Flow sheet of niacin manufacture.

amide Niacin, the least expensive vitamin, is used in foods, feeds, and pharmaceutical preparations to supplement limited amounts available naturally.

A major outlet for niacin is swine and poultry feeds, especially those based on corn. In human nutrition, enrichment of flour and bread with niacin is required in most states. Rice and corn products, alimentary pastes, and milk are fortified with niacin in some areas. Both niacin and nicotinamide are used in single and multivitamin capsules and tablets, and are prescribed for various clinical applications. [A.H.C.]

Nibbling

The cutting of material by the action of a reciprocating punch. The nibbler takes repeated small bites as the work is passed beneath it. The workpiece must be backed up by a support or die. Ferrous and nonferrous metals as well as some nonmetallic compositions may be cut by nibbling. Cuts may be made in mild steel up to approximately ½ in thick.

Nibbling machines are constructed with considerable distance or throat between the punch and its supporting upright. This distance, plus the use of a round punch which allows the workpiece to be moved about, permits the cutting of irregular shapes. Duplicate pieces may be made by using templates as guides for the punch. Tubing may also be cut. Internal holes must be started from previously made holes. *See* MACHINING OPERATIONS. [A.T.]

Nibbling machine being used to cut a cam. (Wilson Mechanical Instrument Division of American Chain and Cable Co.)

Niccolite

A minor ore of nickel. Niccolite is a mineral having composition NiAs and crystallizing in the hexagonal system. Crystals are rare and it usually occurs in massive aggregates with metallic luster and pale copper-red color. Because of the color, not the composition, it is called copper nickel. The hardness is 5.5 (Mohs scale) and the specific gravity is 7.78. Niccolite is frequently associated with other nickel arsenides and sulfides in massive [illegible]. It is [illegible] ver[illegible] ma[illegible] *See* NICKEL; PYRRHOTITE. [C.S.HU.]

Nickel

Chemical element number 28, nickel, Ni, is a silver-gray, ductile, malleable, tough metal. The atomic weight of naturally occurring nickel is 58.71. Iron, cobalt, and nickel are all members of group VIIIb of the periodic table of elements, and these three metals have many chemical similar[illegible] [illegible] but their physical and

chemical properties are mostly distinct from those of iron, cobalt, and nickel.

Nickel consists of five natural isotopes having atomic masses of 58 (68% of natural nickel), 60 (26%), 61 (1%), 62 (4%), and 64 (1%). Six radioactive isotopes have also been identified, having mass numbers of 56, 57, 59, 63, 65, and 66.

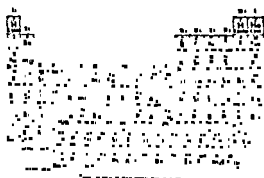
Uses of the metal. Nickel alloys use most of the nickel of commerce. Some of the metal is used in plating to give hard, tarnish-resistant, polishable surfaces. Finely divided nickel is used as a catalyst. Nickel is also used in coinage, in fabrication of special chemical equipment, and in the preparation of nickel compounds. Metallic nickel is sometimes used in the European nickel-cadmium storage battery.

Occurrence. Nickel is a fairly plentiful element, comprising about 0.01% of the igneous rocks. Appreciable quantities of nickel are present in some kinds of meteorite, and large quantities are thought to exist in the earth's core. The two commercially important nickel minerals are pentlandite $(Ni,Fe)_xS_y$, found principally in Ontario, and garnierite $(Ni,Mg)SiO_3 \cdot nH_2O$, found principally in New Caledonia. Other nickel minerals, of less importance, are the red nickel ore (NiAs), millerite or yellow nickel ore (NiS), breithauptite (NiSb), niccolite or white nickel ore ($NiAs_2$), gersdorffite (NiAsS), and ullmannite (NiSbS). In addition to being very important in Canada, sulfide ores are fairly widely distributed in Europe and Asia, and some are present in South Africa and the United States. Some useful nickel silicate ore is found in Cuba. For the most part, the arsenide and antimonide minerals are not important enough to be classed as commercial ores. As with many other metals, the true chemical composition of some of these nickel minerals is not definitely known.

Compounds of nickel

Name	Formula	Uses	Properties and remarks
Nickel sulfate 6-hydrate	$NiSO_4 \cdot 6H_2O$	In nickel plating; in dip baths for enameling, preparation of nickel compounds and catalytic nickel	Green or blue soluble compound, commercially most important of nickel compounds
Nickel chloride 6-hydrate	$NiCl_2 \cdot 6H_2O$	Reagent; in electrorefining of catalytic nickel	Bright green soluble compound
Nickel nitrate 6-hydrate	$Ni(NO_3)_2 \cdot 6H_2O$	Reagent, preparation of nickel compounds and catalytic nickel	Emerald green, very soluble, also exists as other hydrates
Nickel(II) oxide	NiO	In production of alloys, in enamel frits and ceramic glazes, in glass manufacture	Green or black; green form insoluble in water, soluble in acids, black form insoluble in water and acids
Nickel dioxide	NiO_2	Oxidizing agent in Edison storage battery	Black, insoluble
Nickel ammonium sulfate 6-hydrate	$Ni(NH_4)_2(SO_4)_2 \cdot 6H_2O$	Sometimes used in nickel plating	Blue-green soluble compound
Nickel tetracarbonyl	$Ni(CO)_4$	Catalyst, source of carbon monoxide in organic synthesis, source of very pure nickel by decomposition	Colorless, volatile compound, made by reaction of nickel metal with carbon monoxide, contains nickel in oxidation state 0, more poisonous than carbon monoxide
Nickel dimethylglyoxime	$Ni(C_4H_7N_2O_2)_2$	Analytical determination of nickel	Red, insoluble complex compound
Nickel formate 2-hydrate	$Ni(CHO_2)_2 \cdot 2H_2O$	Decomposes at 200–250°C to yield catalytic nickel	Green, moderately soluble compound
Nickel acetate 4-hydrate	$Ni(C_2H_3O_2)_2 \cdot 4H_2O$	Sealer for anodized aluminum, mordant in textile dyeing, reagent in dye preparation	Blue-green soluble compound
Nickel sulfide	NiS	Analytical determination of nickel	Brown or black insoluble compound

Nickel occurs in small quantities (0.1–3 ppm dry weight) in plants and animals. For a discussion of nickel production, *see* NICKEL METALLURGY.



Nickel metal. Nickel is of moderate strength [illegible]

[illegible] particularly significant in the catalytic use of nickel. The density of nickel is 8.90 times that of water at 20°C. 1 ft³ of nickel would weigh about 550 lb. Nickel melts at 1453°C and boils at about 2800°C. Its electrical conductivity is 15% that of copper and 14% that of silver, its heat conductivity is 15% that of silver. Nickel has a high enough value of magnetic susceptibility to be classed as *ferromagnetic*, but it does not equal iron in this respect. Commercial wrought nickel is 99.4% pure. *See* NICKEL ALLOYS.

Chemical properties. Nickel is only moderately reactive. It resists alkaline corrosion and does not burn in the massive state, although fine nickel wires can be ignited. Specially prepared nickel, consisting of very small, porous particles (pyrophoric nickel), burns spontaneously when exposed to the air. Nickel is above hydrogen in the electrochemical series, and it dissolves slowly in dilute acids, releasing hydrogen and forming the green dipositive nickel ion, Ni^{++}. With solutions of oxidizing agents including strong nitric acid, nickel becomes passive and resists attack. In metallic form nickel is a moderately strong reducing agent. Through physicochemical processes, nickel can take up considerable amounts of hydrogen, as can palladium and platinum, especially when the metals are in the finely divided state. Release of the hydrogen to other substances for chemical reaction is one reason for the catalytic action of these metals. Most compounds of nickel are green or blue. *See* HYDROGENATION.

Nickel compounds. Nickel is usually dipositive in its compounds, but it can exist in the oxidation states 0, 1+, 3+, and 4+ as well. Besides the simple nickel compounds, or salts, nickel forms a variety of coordination compounds or complexes. The nickel ion present in water solutions of simple nickel compounds is itself a complex, $[Ni(H_2O)_6]^{++}$. The compounds of dipositive nickel closely resemble those of dipositive cobalt, so that chemical separation of the two metals is often difficult. Some of the more important simple and complex compounds of nickel are listed in the table.

Analytical methods. Nickel can be identified qualitatively by precipitation of either the hydroxide or the sulfide. The hydroxide is not soluble in excess of hydroxide ion and is only slowly soluble in dilute hydrochloric acid. If ammonium hydroxide is the reagent used, the precipitate is soluble in excess because of the formation of the blue $[Ni(NH_3)_6]^{++}$ complex ion. The sulfide can be precipitated in brown or black form by ammonium sulfide or hydrogen sulfide in neutral or alkaline solution. Cobalt undergoes similar reactions; it may be separated from nickel in acetate buffer solution by precipitating the nickel with dimethylglyoxime. The red nickel dimethylglyoxime may be dried and weighed for quantitative determination.

Other quantitative methods include precipitation of the hydroxide, which is heated and weighed as NiO; precipitation with Grossmann's reagent, dicyandiamidine sulfate; and electrolysis followed by weighing of the nickel as the metal. [W.T.C.]

Bibliography: J. C. Bailar, Jr. (ed.), *The Chemistry of the Coordination Compounds*, 1956; J. Kleinberg (ed.), *Treatise on Inorganic Chemistry*, vol. 2, 1956; N. V. Sidgwick, *The Chemical Elements and their Compounds*, vol. 2, 1950.

Nickel alloys

Combinations of nickel with other metals. Nickel has been used in electroplating since 1843 and as an alloying addition to steels since about 1889. It was first used as a base for alloys with the introduction of Monel nickel-copper alloy in about 1905. The nominal compositions of some of the currently available alloys containing more than 50% nickel are given in Table 1.

Nickel base alloys may be melted in open-hearth, electric-arc, or induction furnaces in air, under inert gas, or in vacuum. Casting may also be done under these same ambient conditions. Cast shapes are made in sand or investment molds or by shell molding. Ingots for wrought products are cast in metal molds and are hot worked by forging, rolling, or extruding. In some instances, further work may be done cold by rolling or drawing. Nickel base alloys, made available in this way in bar, rod, wire, plate, strip, sheet, and tubular forms, may be fabricated into finished products using conventional metal working and joining techniques.

Alloyed nickels. D nickel and Duranickel, age-hardenable nickel, are essentially binary alloys with 4.75% manganese and 4.5% aluminum, respectively. Manganese in the first of these, extends the range of applicability in the presence of sulfur by about 300°F to a limiting temperature in the neighborhood of 1000°F. A characteristic use of this material is as wire for spark-plug electrodes.

Aluminum and titanium confer [illegible]

[illegible] and heat treatment. In

chemical properties are mostly distinct from those of iron, cobalt, and nickel.

Nickel consists of five natural isotopes having atomic masses of 58 (68% of natural nickel), 60 (26%), 61 (1%), 62 (4%), and 64 (1%). Six radioactive isotopes have also been identified, having mass numbers of 56, 57, 59, 63, 65, and 66

Uses of the metal. Nickel alloys use most of the nickel of commerce Some of the metal is used in plating to give hard, tarnish-resistant, polishable surfaces. Finely divided nickel is used as a catalyst. Nickel is also used in coinage, in fabrication of special chemical equipment, and in the preparation of nickel compounds Metallic nickel is sometimes used in the European nickel-cadmium storage battery.

Occurrence. Nickel is a fairly plentiful element, comprising about 0.01% of the igneous rocks Appreciable quantities of nickel are present in some kinds of meteorite, and large quantities are thought to exist in the earth's core. The two commercially important nickel minerals are pentlandite $(Ni,Fe)_xS_y$, found principally in Ontario, and garnierite $(Ni,Mg)SiO_3 \cdot nH_2O$, found principally in New Caledonia Other nickel minerals, of less importance, are the red nickel ore (NiAs), millerite or yellow nickel ore (NiS), breithauptite (NiSb), niccolite or white nickel ore ($NiAs_2$), gersdorffite (NiAsS), and ullmannite (NiSbS). In addition to being very important in Canada, sulfide ores are fairly widely distributed in Europe and Asia, and some are present in South Africa and the United States. Some useful nickel silicate ore is found in Cuba For the most part, the arsenide and antimonide minerals are not important enough to be classed as commercial ores. As with many other metals, the true chemical composition of some of these nickel minerals is not definitely known

Compounds of nickel

Name	Formula	Uses	Properties and remarks
Nickel sulfate 6-hydrate	$NiSO_4 \cdot 6H_2O$	In nickel plating, in dip baths for enameling; preparation of nickel compounds and catalytic nickel	Green or blue soluble compound, commercially most important of nickel compounds
Nickel chloride 6-hydrate	$NiCl_2 \cdot 6H_2O$	Reagent, in electrorefining of catalytic nickel	Bright green soluble compound
Nickel nitrate 6-hydrate	$Ni(NO_3)_2 \cdot 6H_2O$	Reagent, preparation of nickel compounds and catalytic nickel	Emerald green, very soluble; also exists as other hydrates
Nickel(II) oxide	NiO	In production of alloys, in enamel frits and ceramic glazes, in glass manufacture	Green or black; green form insoluble in water, soluble in acids, black form insoluble in water and acids
Nickel dioxide	NiO_2	Oxidizing agent, in Edison storage battery	Black, insoluble
Nickel ammonium sulfate 6-hydrate	$Ni(NH_4)_2(SO_4)_2 \cdot 6H_2O$	Sometimes used in nickel plating	Blue-green soluble compound
Nickel tetracarbonyl	$Ni(CO)_4$	Catalyst; source of carbon monoxide in organic synthesis, source of very pure nickel by decomposition	Colorless, volatile compound; made by reaction of nickel metal with carbon monoxide, contains nickel in oxidation state 0, more poisonous than carbon monoxide
Nickel dimethylglyoxime	$Ni(C_4H_7N_2O_2)_2$	Analytical determination of nickel	Red, insoluble complex compound
Nickel formate 2-hydrate	$Ni(CHO_2)_2 \cdot 2H_2O$	Decomposes at 200-250°C to yield catalytic nickel	Green, moderately soluble compound
Nickel acetate 4-hydrate	$Ni(C_2H_3O_2)_2 \cdot 4H_2O$	Sealer for anodized aluminum, mordant in textile dyeing, reagent in dye preparation	Blue-green soluble compound
Nickel sulfide	NiS	Analytical determination of nickel	Brown or black insoluble compound

high strength and oxidation resistance, have found application in jet engine and gas turbines for such parts as combustion liners, blades, vanes, and disks.

In the interest of optimizing properties, these complex nickel chromium alloys are being produced in increasing quantities by vacuum-melting and vacuum-pouring techniques.

A cast heat-resisting alloy carrying the Alloy Castings Institute designation HW (nominally 60% nickel, 12% chromium, 23% iron) is used principally for furnace parts and heat-treating fixtures. It has good resistance to oxidation and carburization, only modest hot strength, but good thermal shock resistance

Hastelloy alloys. Hastelloy alloys B, C, and D are used primarily in corrosive environments. Hastelloy B is resistant to hydrochloric and sulfuric acids within certain limits of concentration, temperature, and degree of aeration It is not recommended for service involving strong oxidizing acids or oxidizing salts Hastelloy C is unusually resistant to oxidizing solutions and to moist chlorine. Hastelloy D has exceptional resistance to hot concentrated sulfuric acid.

Nickel-iron alloys. Alloys containing more than 50% nickel are used in various applications involving controlled thermal expansivity or certain magnetic requirements In the range 50-52% nickel, the balance iron, the alloys have thermal expansion characteristics useful in making some types of glass-to-metal seals

In the range 77-80% nickel, with or without about 4% molybdenum, the balance iron, the alloys have very high initial and maximum magnetic permeabilities when properly processed. See ALLOY; IRON ALLOY, NICKEL, NICKEL METALLURGY, STAINLESS STEEL [R J R]

Nickel metallurgy

The [illegible]

[illegible] silver, was used in the sixteenth century in China, and early weapons were often fashioned [illegible]

A [illegible]

17[illegible] [illegible] it in relatively pure form.

Occurrence. Although nickel ranks [illegible] fou[illegible] ign[illegible] are [illegible] of commercial importance Nickel ores are of two generic types, sulfides and laterites

Sulfides In ores of this type, nickel is present chiefly as the mineral pentlandite, a nickel-iron sulfide, usually in association with pyrrhotite and chalcopyrite The most important known deposits, at Sudbury, Canada, have provided the major portion of the world's nickel supply since 1905 Other substantial deposits have been developed in the Thompson-Moak Lake and in the Lynn Lake areas of northern Manitoba, Canada. Russia exploits deposits on the Kola peninsula and near Norilsk in Siberia

Laterites. Lateritic nickel ores occur as oxide ores, in which the nickel is dispersed through limonite, and silicate ores, in which the nickel occurs in a hydrated magnesium silicate. Lateritic ores are widely distributed throughout the tropics and constitute the world's largest known reserves of nickel Deposits in Cuba, New Caledonia, and Oregon are being worked commercially, and there are extensive reserves in Indonesia, the Philippines, Latin America, and the Soviet Union.

Production and uses. Nickel gained commercial prominence late in the nineteenth century when substantial reserves were found in New Caledonia and at Sudbury, and the world's naval powers adopted nickel-bearing armor. Until about 1920, nickel markets hinged upon military requirements. Following World War I, research into industrial applications was greatly increased and the success of this continuing program is evident from nickel's many diversified and expanding uses

Nickel is marketed in various forms. Its price is based on electrolytic nickel which has remained at 74¢/lb since December, 1956. During 1957, nickel consumption in the United States was distributed as shown in Table 2.

Extractive metallurgy. Selection of processes for nickel extraction is largely determined by the type of ore to be treated Sulfide ores are amenable to concentration by such methods as flotation or magnetic separation The state of combination of nickel

Table 1. Nickel production, short tons

Estimated total world nickel since 1870, annual		Estimated world mine production in 1957	
Year	Production	Country	Production, contained nickel
1870	500	Canada	189,000
1900	10,200	U S S R	55,000
1918	52,500	New Caledonia	33,000
1940	154,300	Cuba	22,000
1957	314,000	United States	10,000
		Other	5,000
		Total	314,000

Table 2. Nickel consumption in the United States

Use	Per cent of total*
Nonferrous alloys	27 3
Stainless steels	22 0
Electroplating	20 0
Other alloy steels	13 0
High-temperature and electrical-resistance alloys	
Cast iron	8 0
Catalysts, magnets, ceramics, other	4 5
	5 2

* From U S Bureau of Mines, *Minerals Yearbook*, and other U S Government publications

Table 1. Nominal composition of some nickel-base alloys, weight per cent

Trademark	Ni	Cu	Cr	Co	Mo	Ti	Al	Cb	Fe	Mn	Si	C	Other
D	95									4.75		0.08	
Duranickel	93.7	0.05				0.4	4.4		0.35	0.3	0.5	0.17	
Monel	66	31.5							1.35	0.9	0.15	0.18	
K Monel	66	29				0.5	2.75		0.9	0.75	0.5	0.15	
S Monel	63	30							2	0.75	4	0.1	
Chromel P	90		10										
Nichrome V	80		19.5							2.5*	1	0.25	
Alumel	94						2			3	1		
Nimonic 75	Bal	0.5*	19.5			0.4			5*	1*	1*	0.12	
Nimonic 80A	Bal		19.5	2*		2.2	1.1		5*	1*	1*	0.1*	
Inconel	Bal	0.5*	15.5						8	1*	0.5*	0.15	
Inconel X	Bal	0.2*	15			2.5	0.9	0.9	7	0.7	0.4	0.04	
Inconel 713C	Bal		12		4	0.5	6	2	5*	1*	1*	0.2*	
Udimet 500	Bal		17.5	16.5	4	2.9	2.9		4*	0.75*	0.75*	0.15*	
Waspaloy	Bal		19	14	3	2.5	1.2		2	0.7	0.4	0.05	
M252	55		19	10	10	2.5	0.75		2	1	0.7	0.1	
GMR 235	Bal		15.5		5	2.5	3		10	0.25*	0.6*	0.15	0.06 B
Hastelloy B	61		1*	2.5*	27.5	2			5.5	1*	1	0.05*	0.4 V
Hastelloy C	54		15.5	2.5*	15.5				5.5	1*	1*	0.08*	0.35 V*, 4W
Hastelloy D	82	3	1*	1.5*					2*	1	9	0.12*	

* Maximum

this condition, it is well suited to the manufacture of springs and diaphragms.

Monel nickel-copper alloy. This alloy contains about two-thirds nickel and one-third copper and is the oldest of the commercial nickel-base alloys, dating from about 1905 when it was directly smelted from the copper-nickel matte obtained from Sudbury, Ontario, sulfide ore. The good fabricating characteristics and corrosion-resistance of this alloy have made it widely used in marine applications and in the chemical-processing and petroleum industries. It has applicability in the new and expanding field of nuclear propulsion. As with nickel, this alloy can be made age-hardenable by the addition of aluminum and titanium K Monel age-hardenable nickel-copper alloy has corrosion-resisting characteristics similar to the nonage-hardenable composition, and is widely specified for such applications as marine propellers, shafting, valves, pump parts, and springs A usable tensile strength of about 175,000 psi is obtainable in cold drawn and age-hardened wire S Monel hard nickel-copper cast alloy, which is age-hardenable by virtue of its relatively high silicon content (4.0%), possesses nongalling and antiseizing characteristics which make it applicable to ball- and roller-bearing races.

Nickel-chromium alloys. Nickel-chromium binary alloys are used primarily in specialty high-temperature service. Nichrome V is a common high-quality resistance-heating-element material possessing good resistance to oxidation up to about 2100°F, superior to either of its two component elements. The alloy is used both in industrial-furnace and household-appliance heating elements Chromel P is used with the chromium-free nickel-base Alumel in temperature-sensing devices known as thermocouples. This particular alloy couple has favorable thermoelectric characteristics for applicability in the measurement of temperatures up to 2000°F.

Nickel-chromium and related complex alloys are widely used for structural and general-purpose applications at high temperatures and in certain corrosive environments, particularly where freedom from stress-corrosion cracking is essential. In this latter instance, Inconel nickel-chromium alloy has applicability in nuclear-propulsion units.

This class of alloys encompasses a broad range of high temperature properties Nimonic 75 nickel-chromium alloy is widely used as a scale-resistant sheet material Neither this material nor Inconel nickel-chromium alloy responds to age hardening, and hence they are on the low side of the elevated-temperature mechanical property range By contrast, Inconel X age-hardenable nickel-chromium alloy, which contains added aluminum, titanium, and columbium, develops greatly improved high-temperature strength after proper heat treatment. The more highly alloyed Inconel 713C nickel-chromium cast alloy exhibits further strength improvement at the high side of the temperature range of applicability for the complex nickel-chromium alloys. For comparison, 100-hour rupture strengths of these three alloys are listed in Table 2.

The aforementioned materials and a number of similar proprietary alloys such as Udimet 500, M252, Waspaloy, and GMR235, which combine

Table 2. Rupture strengths (100 hours) of nickel-chromium alloys, psi

Alloy	1500°F	1700°F
Inconel	7,000	3,000
Inconel X	25,000–30,000	8,000–10,000
Inconel 713C	58,000	30,000

contains most of the precious metals After crushing and grinding, the metallics are removed magnetically for recovery of metal values, and the main sulfide fraction is separated into copper sulfide and nickel sulfide concentrates by froth flotation See FLOTATION.

Most of the nickel sulfide concentrate is converted to granular nickel oxide on Dwight Lloyd sintering machines A portion of this product is marketed directly for alloy steel production. Two processes are employed to convert the remaining oxide to pure metal for market. One involves reduction smelting to metal anodes followed by electrolytic refining, using a sulfate-chloride electrolyte with divided cells and continuous electrolyte purification. By-products of this operation are cobalt and precious metals The other is the Mond process, described below. Part of the nickel sulfide concentrate from slow-cooled matte is melted, cast directly into sulfide anodes, and electrolytically refined to nickel cathodes, with concurrent recovery of elemental sulfur, cobalt, selenium, and precious metals.

The Mond nickel carbonyl process is used in Great Britain to produce refined nickel from nickel oxide sinter. This material is ground, calcined, reduced with water gas, and treated with carbon monoxide at about 120°F to volatilize nickel as nickel carbonyl. This compound is decomposed at about 400°F to yield high purity nickel in pellet form Copper and cobalt salts and precious metals are recovered from the residue. Nickel powder is also produced at this plant in a pressure carbonyl system.

At Falconbridge, the bulk flotation concentrate is sintered and blast-furnace smelted to a copper-nickel matte, following conventional copper smelting practice. The Bessemer matte is shipped to a refinery in Norway, where it is ground, roasted to oxide, and leached with dilute sulfuric acid to remove copper, which is recovered electrolytically from the leach solution The nickel bearing residue

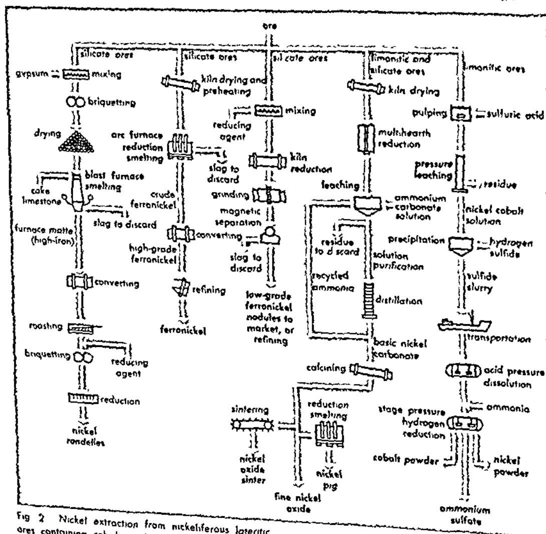


Fig 2 Nickel extraction from nickeliferous lateritic ores containing cobalt, and varying amounts of iron and gangue

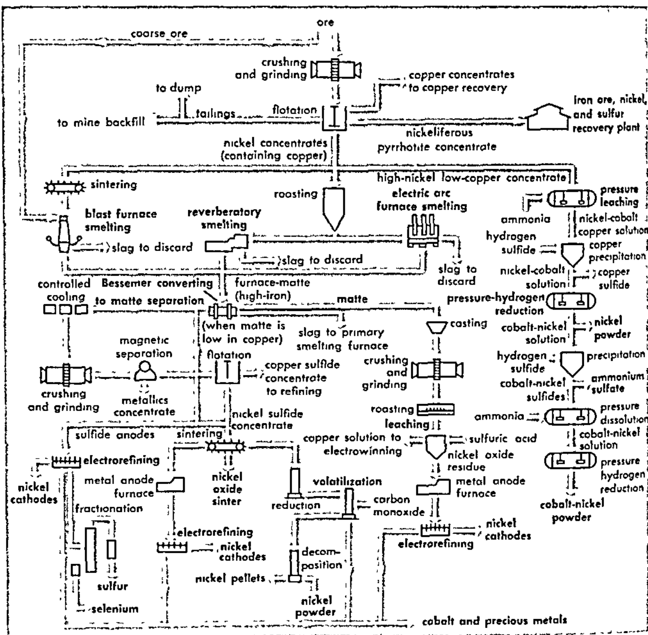


Fig. 1. Nickel extraction from nickeliferous sulfide ores containing copper, nickel, cobalt, iron, sulfur, precious metals, and gangue.

in the lateritic ores usually precludes such enrichment, thus requiring treatment of the total ore.

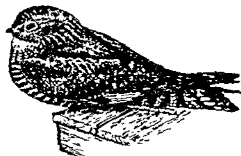
Sulfide ores. Figure 1 illustrates graphically the extractive metallurgy of sulfide ores. A minor amount of higher-grade ore is smelted directly in blast furnaces. With lower-grade ores, averaging up to 2.5% nickel plus copper, mineral values are liberated by crushing and grinding, and separated from the gangue by froth flotation. *See* ORE DRESSING.

International Nickel Company employs selective flotation and magnetic separation to divide the bulk concentrate into nickel-, copper-, and iron-rich fractions for separate treatment. A high-grade iron ore, nickel, and sulfuric acid are recovered from the iron concentrate. The nickel concentrate is treated by pyrometallurgical processes. The major portion undergoes partial roasting in Herreshoff furnaces to eliminate about half the sulfur and to oxidize the associated iron. The hot calcine, plus flux, is smelted in coal fired reverberatory furnaces operating at about 2200°F Other procedures involve sintering in Dwight-Lloyd machines followed by blast-furnace smelting, or partial roasting followed by electric-arc-furnace smelting. The product of these operations, termed furnace matte, is transferred to Pierce-Smith converters and blown with air to oxidize the remaining iron and associated sulfur, yielding Bessemer matte containing nickel, copper, cobalt, small amounts of precious metals, and about 22% sulfur

The molten Bessemer matte is cast into 25-ton molds in which it undergoes controlled slow cooling to promote formation of relatively large, discrete crystals of copper sulfide, nickel sulfide, and a small quantity of a metallic phase which

Nighthawk

A bird, *Chordeiles minor*, frequently called the bull bat, a member of the Caprimulgidae. It is quite similar to the whippoorwill, but can readily be distinguished from the latter by the large white patch in the middle of each wing. It is common over most of the United States and southern Canada. The nighthawk has accepted civilization and is a common city resident, nesting on the flat topped roofs of large buildings as readily as on the forest



The nighthawk, *Chordeiles minor*; length to 10 in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

floor or prairie sod. During its acrobatic hunt for flying insects, the nighthawk makes frequent spectacular dives and gives voice to a loud cry. *See* CAPRIMULGIFORMES, WHIPPOORWILL. [J D B.]

Niobium

Chemical element number 41, *niobium*, Nb, has atomic weight 92.91. The name first used for this element in the United States was columbium. Since 1951 American chemists have adopted the name niobium in conformance with the recommendation of the International Union of Pure and ...

Niobium is a member of the fifth subgroup of the periodic table and is in the 4*d* transitional series. Its valence electron configuration is $4d^45s^1$ which results in a maximum oxidation state of V. Oxidation states of II, III, and IV are also reported. The calculated radius for the Nb^{5+} ion is 0.70 Å. Niobium is found in nature with its homolog, tantalum, although it occurs by itself in pyrochlore, $NaCaNb_2O_6$. Niobium occurs in the earth's crust to the extent of 2.4×10^{-3}% which makes it about 10 times as abundant as tantalum. For a discussion of the metallurgical separation and reduction, *see* TANTALUM.

There are no large-scale uses for niobium metal, although it seems to have good potential as a structural material above 1000°C. It also should be useful in nuclear reactors, because of its low cross section for capture of thermal neutrons. Some of its compounds exhibit superconductivity at relatively high temperatures. Most niobium is used in stainless steel and high-temperature alloys.

Niobium metal. Niobium metal has a density of 8.6 g/cm³ and crystallizes in the body-centered cubic system. It has a melting point of 2468°C. Its cross section for capture of thermal neutrons is only 1.1 barns. Although the standard oxidation potential for the reaction



$$2Nb + 5H_2O \rightarrow Nb_2O_5 + 10H^+ + 10e^-$$

is estimated to be +0.62 volt, the metal is quite inert to all acids except hydrofluoric acid. Niobium is slowly oxidized in alkaline solutions. The halogens and oxygen react with it on heating, and it combines with nitrogen and carbon to form ...

... are the most important binary compounds of niobium. The oxide, Nb_2O_5, has a melting point of 1520°C and is obtained by heating the metal in oxygen. The oxide dissolves in fused alkali to yield complex niobates such as $Nb_6O_{19}^{8-}$ and $Nb_2O_7^{4-}$ which are water soluble and hydrolyze to the insoluble oxide. The normal niobates NbO_4^{3-} are insoluble in water. The freshly precipitated oxide is a very weak acid and dissolves slightly in aqueous alkali. The oxide dissolves in hydrofluoric acid to give ionic species such as $NbOF_5^{2-}$ and $NbOF_6^{4-}$, depending upon the fluoride and hydrogen-ion concentrations. In a very large excess of fluoride ... K_2NbF ... upon the ... tions ... in solutions of organic dibasic acids and hydroxy carboxylic acids.

The niobium pentahalides are all low melting, low boiling compounds which hydrolyze in water to give the oxide. They may be prepared by direct action of the halogen or dry halogen halide on the metal. Preparations based on the oxide and free of oxyhalide impurity are difficult to obtain. High-purity pentachloride can be obtained from the

is reduction smelted and cast to metal anodes for electrolytic refining. Cobalt and precious metals are also recovered.

Sherritt Gordon Mines, Ltd. separates the bulk flotation concentrate into nickel and copper concentrates, and treats the nickel concentrate by a pressure hydrometallurgical process A suspension of relatively high-grade, finely ground nickel concentrate is leached with ammoniacal solution with vigorous aeration, at temperatures up to 175°F and pressures of 100–125 psi. The metal sulfides are oxidized essentially to sulfates, the iron hydrolyzes and precipitates into the residue which is discarded, while nickel, copper, and cobalt dissolve. After treatment to remove copper and trace metals, the ammoniacal nickel-cobalt solution is autoclaved with hydrogen at about 375°F and 450 psi to yield a nickel powder, the bulk of which is marketed in briquette form. The residual solution, with further treatment, yields a nickel-cobalt powder. Ammonium sulfate is a by-product of this operation.

Laterites. The lateritic ores contain practically no copper. Those of New Caledonia now being worked contain about 3% nickel. They are blast-furnace smelted with coke and gypsum to produce nickel matte for converting and subsequent refining. Ore is also smelted to ferronickel in a new electric furnace plant (Fig. 2).

Cuban laterites, which average about 1.5% nickel plus cobalt, are treated by a combined pyro- and hydrometallurgical process in a plant at Nicaro, Cuba. The ore is surface mined, dried, ground, and selectively reduced with producer gas in multihearth furnaces at temperatures rising to about 1300°F. The bulk of the iron is reduced to magnetite, while the nickel is reduced to metal. The reduced ore is countercurrently leached in an aerated ammoniacal ammonium carbonate solution, to dissolve nickel selectively as nickel ammonium carbonate The pregnant solution is distilled in bubble-cap towers to remove ammonia and part of the carbon dioxide for recycle, whereupon the nickel precipitates as basic nickel carbonate The precipitate is calcined to an oxide powder Some is marketed in this form and some after sintering. *See* SINTERING.

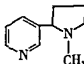
In the Freeport Nickel Company plant, nickel and cobalt are selectively dissolved from Cuban limonitic nickel ore by pressure sulfuric acid leaching, and after iron removal, are precipitated with hydrogen sulfide. Coral is used for pH control A slurry of the precipitated sulfides is shipped to the United States for refining by a modification of the above-described Sherritt Gordon process. Final products are nickel and cobalt powders or briquettes.

Several other operations, in Japan, the United States, Europe, and elsewhere, produce modest amounts of nickel, often as by-products. The Soviet Union has become a substantial producer of nickel. For the most part, the processes used are similar to those already described. *See* NICKEL; NICKEL ALLOYS; PYROMETALLURGY, NONFERROUS.

[L.S RE.]

Nicotine

An alkaloid present in tobacco, the dried leaf of *Nicotiana tabacum*. It was first isolated by L. Vaquelin in 1809 from tobacco smoke and characterized in a pure form. Nicotine has the following structure:



Nicotine is one of the few liquid alkaloids It is colorless, volatile, alkaline in solution, and, on exposure to air, it turns brown and acquires the odor of tobacco. It is readily soluble in water and forms water-soluble salts. It exists in the tobacco leaf in combination with acetic, malic, and citric acids.

Isolation is accomplished by making an aqueous extract of the powdered plant, adding alkali, and steam-distilling the volatile alkaloid from the mixture. Synthesis of the alkaloid was accomplished in 1904. However, the synthesis is of little commercial interest since nicotine may be obtained very economically from natural sources (tobacco refuse). The principal value has been to confirm the postulated structure. Nicotine (as the sulfate) finds wide use as an insecticide. It also serves as a source of nicotinic acid amide (an important B-vitamin).

Because of its toxicity, nicotine is little used medically. The alkaloid is one of the most toxic of all substances and acts with a rapidity comparable to that of cyanide Poisoning has occurred from accidental ingestion of insecticide sprays containing nicotine. Also, the lay use of tobacco infusion as an enema against intestinal parasites has resulted in death. The nicotine content of one cigar approximates the lethal dose for man. However, swallowed in the form of tobacco, nicotine is much less toxic than would be expected. Children have ingested cigarettes without lethal effect, despite the fact that each cigarette contains nearly the estimated fatal dose of nicotine Apparently the gastric absorption of nicotine from tobacco is delayed, so that vomiting caused by the initially absorbed fraction usually removes the tobacco remaining in the stomach.

Although nicotine has no therapeutic uses, it is of great pharmacological interest, and studies with nicotine have contributed greatly to the understanding of the physiology of the nervous system. In addition, the complex respiratory and circulatory changes occurring in the body after administration of nicotine make it an interesting tool in experimental pharmacology. *See* ALKALOID

[S M K.]

Nighthawk

A bird, *Chordeiles minor*, frequently called the bull bat, a member of the Caprimulgidae. It is quite similar to the whippoorwill, but can readily be distinguished from the latter by the large white patch in the middle of each wing. It is common over most of the United States and southern Canada. The nighthawk has accepted civilization and is a common city resident, nesting on the flat-topped roofs of large buildings as readily as on the forest

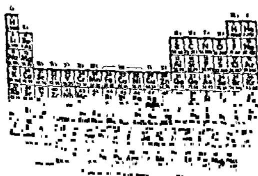


The nighthawk, *Chordeiles minor*; length to 10 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

floor or prairie sod. During its acrobatic hunt for flying insects, the nighthawk makes frequent spectacular dives and gives voice to a loud cry. *See* CAPRIMULGIFORMES; WHIPPOORWILL. [J.D.B.]

Niobium

Chemical element number 41, niobium, Nb, has atomic weight 92.91. The name first used for this element in the United States was columbium. Since 1951, American chemists have adopted the name niobium in conformance with the recommendation



of the International Union of Pure and Applied Chemistry Nomenclature Committee. However, metallurgists and the metals industry in general have been unwilling to make the change and still use the name columbium.

Niobium is a member of the fifth subgroup of the periodic table and is in the 4d transitional series. Its valence-electron configuration is $4d^4 5s^1$ which results in a maximum oxidation state of V. Oxidation states of II, III, and IV are also reported. The calculated radius for the Nb^{5+} ion is 0.70 A. Niobium is found in nature with its homolog, tantalum, although it occurs by itself in pyrochlore, $NaCaNb_2O_6$. Niobium occurs in the earth's crust to the extent of 2.4×10^{-3}% which makes it about 10 times as abundant as tantalum. For a discussion of the metallurgical separation and reduction, *see* TANTALUM.

There are no large-scale uses for niobium metal, although it seems to have good potential as a structural material above 1000°C. It also should be useful in nuclear reactors, because of its low cross section for capture of thermal neutrons. Some of its compounds exhibit superconductivity at relatively high temperatures. Most niobium is used in stainless steel and high-temperature alloys.

Niobium metal. Niobium metal has a density of 8.6 g/cm^3 and crystallizes in the body-centered cubic system. It has a melting point of 2468°C. Its cross section for capture of thermal neutrons is only 1.1 barns. Although the standard oxidation potential for the reaction

$$2Nb + 5H_2O \rightarrow Nb_2O_5 + 10H^+ + 10e^-$$

is estimated to be +0.62 volt, the metal is quite inert to all acids except hydrofluoric acid. Niobium is slowly oxidized in alkaline solutions. The halogens and oxygen ...



...ride, bromide, and iodide of the oxidation state V are the most important binary compounds of niobium. The oxide, Nb_2O_5, has a melting point of 1520°C and is obtained by heating the metal in oxygen. The oxide dissolves in fused alkali to yield complex niobates such as $Nb_6O_{19}^{8-}$ and $Nb_2O_7^{4-}$ which are ...

... weak acid and dissolves slightly in aqueous alkali. The oxide dissolves in hydrofluoric acid to give ionic species such as $NbOF_5^{2-}$ and $NbOF_7^{4-}$, depending upon the fluoride and hydrogen-ion concentrations. In a very large excess of fluoride ion, K_2NbF_7 crystallizes from solution. The oxide also dissolves in concentrated hydrochloric acid to give species which have been identified as $Nb(OH)_2Cl_4^-$, $Nb(OH)_2Cl_3$, and $Nb(OH)Cl^{3+}$, depending upon the hydrogen-ion and chloride-ion concentrations. The oxide also shows solubility in concentrated sulfuric acid and in solutions of organic dibasic acids and hydroxy carboxylic acids.

The niobium pentahalides are all low-melting, low-boiling compounds which hydrolyze in water to give the oxide. They may be prepared ...

... difficult to obtain. High-purity pentachloride can be obtained from the

oxide by reaction with carbon tetrachloride under pressure or by refluxing in hexachlorobutadiene, with which tantalum oxide does not react under the time and temperature conditions used for niobium oxide. This may be a basis for a separation procedure. The reduced states of niobium are slightly better known than those of tantalum. The compound $NbCl_4$ has been prepared by reduction with niobium metal, and the trihalide by the reduction of the pentahalide with hydrogen in a discharge tube. The NbF_3 compound has been prepared by the action of a hydrogen–hydrogen fluoride gas mixture on $NbH_{0.7}$. The $NbCl_3$ and $NbBr_3$ compounds were prepared by hydrogen reduction of the corresponding pentahalides at 400–500°C in a hot-cold tube. The trichlorides and tribromides can be sublimed in vacuum at 400°C unchanged.

In aqueous solutions, niobium(IV) species have been produced by electrolytic reduction. In the absence of complexing ions, the IV state disproportionates to the V and III states. The reduced states undoubtedly are not stable with respect to oxidation by water, except as complex species

Niobium of 90% purity or better can be determined by developing the peroxyniobate color with 30% hydrogen peroxide in a solution containing the sample dissolved in concentrated sulfuric and phosphoric acids, under which conditions the peroxytantalate color does not appear. The absorbency is measured at 325 mμ. [E.M.L.]

Bibliography: C. A. Hampel (ed.), *Rare Metals Handbook*, 1954

Nippotaeniidea

An order of tapeworms of the subclass Cestoda The few known species are intestinal parasites of Eurasian fresh-water fishes. The head bears a single terminal sucker (see illustration). The segmental



Scolex of *Nippotaenia*

anatomy shows relationships to the Pseudophyllidea and Cyclophyllidea. The life history is unknown. It is probable that this order is related to the proteocephalids. *See* CESTODA; *see also* CYCLOPHYLLIDEA; PSEUDOPHYLLIDEA [C.P.R.]

Niter

A potassium nitrate mineral with chemical composition KNO_3. Niter crystallizes in the orthorhombic system, generally in thin crusts and delicate acicular crystals; it also occurs in massive, granular, or earthy forms. It has good cleavage in three directions; fracture is subconchoidal to uneven; it is brittle; hardness is 2 on Mohs scale; specific gravity is 2.109; the luster is vitreous; and the color and streak are colorless to white *See* NITRATE MINERALS.

Niter is commonly found, usually in small amounts, as a surface efflorescence in arid regions and in caves and other sheltered places. It is usually associated with soda-niter, epsomite, nitrocalcite, and gypsum. The mineral may occur as an efflorescence on soils rich in organic matter from the action of certain bacteria on nitrogenous or animal matter.

Niter occurs associated with soda-niter in the desert regions of northern Chile, and in various other similar occurrences in Italy, Egypt, U.S.S.R., western United States, and elsewhere. It was formerly found in some abundance in limestone caves in Tennessee, Kentucky, Alabama, and Ohio, and was used for the manufacture of gunpowder during the War of 1812 and the Civil War. [C.S.]

Nitrate

The negative ion, NO_3^-, derived from nitric acid, HNO_3. Because almost all metallic nitrates are water-soluble, they are not found in nature but are produced from nitric acid One notable exception to this is the impure sodium nitrate, called Chile saltpeter, which occurs in large amounts because of the aridity and very small rainfall of the Chilean coastal plain.

Because nitrates contain nitrogen in its highest oxidation state (5+), the ion is a useful oxidizing agent. Because of this property, nitrates are often constituents of matches and explosives Ammonium nitrate will detonate when subjected to shock according to the following equation:



$$NH_4NO_3 \rightarrow N_2O + 2H_2O$$

Nitrates are an important source of nitrogen in fertilizers

The brown-ring test is a common qualitative test for the nitrate ion A brown ring forms at the juncture of a dilute ferrous sulfate solution layered on top of concentrated sulfuric acid if the upper layer contains nitrate ion If nitrite ion is present, the brown color appears throughout the solution The brown color is due to the complex ion, $Fe(NO)^{2+}$.

Organic compounds containing the $-NO_2$ group are called nitro compounds and include explosives such as trinitrotoluene (TNT) *See* EXPLOSION AND EXPLOSIVE; FERTILIZER; NITRIC ACID; NITRITE; NITROGEN [F.E.W.R.]

Nitrate minerals

Nitrate minerals are few in number, and with the exception of soda-niter are of rare occurrence Normal anhydrous and hydrated nitrates occurring as minerals are soda niter, $NaNO_3$, niter, KNO_3; ammonia niter, NH_4NO_3, nitrobarite, $Ba(NO_3)_2$, nitrocalcite, $Ca(NO_3)_2 \cdot 4H_2O$, and nitromagnesite, $Mg(NO_3)_2 \cdot 6H_2O$ In addition there are three known naturally occurring nitrates containing hydroxyl or halogen, or compound nitrates. They are gerhardtite, $Cu_2(NO_3)(OH)$ ·

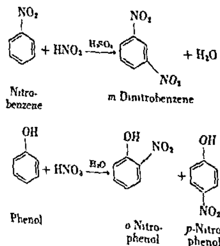
buttgenbachite, $Cu_{19}(NO_3)_2Cl_4(OH)_{32}\cdot 3H_2O$; and darapskite, $Na_3(NO_3)(SO_4)\cdot H_2O$. *See* NITER; SODA NITER

The natural nitrates are, for the most part, readily soluble in water. For this reason they occur most abundantly in arid regions, particularly in South America along the Chilean coast. *See* FERTILIZER, NITROGEN. [C.S.]

Nitration

A substitution reaction in which a nitro group ($-NO_2$) is introduced into a molecule or ion in *place of* hydrogen or some other atom or group

Nitration of aromatic compounds, brought about by treating an aromatic compound with nitric acid in a suitable solvent, is the most common type of nitration Nitrations are electrophilic aromatic substitution reactions, and are subject to the usual directing and activating (or deactivating) effects of substituents The following are typical nitrations:



Nitration reactions are of great practical importance Many explosives are nitro compounds, an example is 2 4 6-trinitrotoluene (TNT), formed by the nitration of toluene Other nitro compounds are useful chemical intermediates, for example, appropriate reduction of nitro compounds gives amino compounds which are of value in the preparation of dyes, pharmaceuticals, polymers, and other end products

The usual mechanism of nitration involves the production of nitronium ion (NO_2^+) from the nitric acid, and then the attack of nitronium ion on the compound being nitrated Sulfuric acid a common solvent for nitrations, brings about 100% conversion of nitric acid to nitronium ion In other solvents, such as acetic acid or nitromethane, or in straight nitric acid, a relatively small fraction of the nitric acid is present as nitronium ion at any one time, but more nitronium ion is formed as this ion is consumed by the nitration reaction. Strongly activated compounds, such as phenols and aromatic amines can be nitrated in aqueous media in which nitronium ion cannot exist in significant amount Such nitrations are catalyzed by nitrous acid; the mechanism is one of nitrosation (formation of a nitroso compound) followed by oxidation of the nitroso compound to a nitro compound The oxidizing agent is nitric acid, which is reduced to regenerate nitrous acid.

Nitration of paraffin hydrocarbons is accomplished by the gas phase reaction of paraffins with nitric acid at 420°C. A complex mixture of products is formed, representing introduction of nitro groups at all possible positions, plus fragmentation of the original paraffin molecule with production of nitro derivatives of lower paraffins. Thus nitration of propane gives 1- and 2-nitropropane, plus nitroethane and nitromethane. This gas-phase nitration has a free-radical mechanism. *See* SUBSTITUTION REACTION. [J.F.B.]

Nitration of cyclic compounds. Nitration of aromatic compounds is generally conducted as a liquid phase reaction, and the organic and acid phases are immiscible so that agitation is necessary to provide adequate contact between the reactants. Groups already substituted on the ring affect the electron distribution, which, together with steric considerations, affects the ease of nitration as well as the specific carbon of the ring at which the nitro group attaches itself Toluene, for example, is nitrated much more easily than benzene, but the introduction of one nitro group on the ring hinders further nitration. Nitration reactions are always highly exothermic, and about 15–35 kcal/mole of heat is evolved

Nitrobenzene, often employed in the manufacture of aniline and as a solvent, is formed in greater than 99% yield by the nitration of benzene using mixed acids, for example, 39% HNO_3 and 55% H_2SO_4. Good heat transfer must be provided when such strong acids are employed. Sulfuric acid has in the past been considered to be a dehydrating agent for removal of the water formed during reaction and to allow the reaction to go to completion This interpretation is incorrect because the reaction is irreversible (the free energy change is large and negative) Instead the sulfuric acid acts as a catalyst. Wh[illegible] it reacts with [illegible] NO_2^+ These [illegible] the case of the [illegible] di- and [illegible] may n[illegible] the mc[illegible] this l[illegible] [illegible] certainty Perchloric acid, acetic anhydride, hydrogen fluoride, and boron trifluoride can be used instead of sulfuric acid in mixed acids to produce nitronium ions.

Dinitrobenzene is manufactured for reduction to *m*-nitroaniline or *m*-phenylenediamine Various chloronitrobenzenes, *p*-nitroacetanilides, and nitrotoluenes, as well as nitronaphthalenes, are employed as intermediates They generally require stronger mixed acids and give lower conversions

and yields than does nitrobenzene because of partial oxidation.

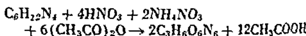
Trinitrotoluene (TNT) is a powerful explosive which is usually produced in a three-step nitration in order to secure economy of the nitrating acids, to reduce oxidation, and to secure best yields with only two nitrators necessary. The first nitration is carried out in the so-called mono house at 140°F, introducing toluene under agitation and with cooling into the cycle acid, followed by mono-mixed acid. The use of the cycle acid (last of the spent acid from previous batch) saves some nitric acid and increases the initial volume to secure better agitation and cooling. The resulting nitration product, so called mono oil, is conducted to the bi-tri house where binitration is performed at 90–180°F using bi-mixed acid. The bi-spent acid is settled, drawn off, and fortified to make mono-mixed acid. The binitration product is nitrated in the same kettle to TNT starting at 180° and ending at 230°F by the use of tri-mixed acid, following which the tri-spent acid is drawn off and sent to the fortifier for further use. The hot tri oil is withdrawn from the nitrator, crystallized in water, and washed in suspension with dilute soda ash solution followed by acidulated 16% Na_2SO_3 (Sellite) to remove undesirable isomers. After a water wash, the TNT is melted, washed, solidified or gained, and flaked or filled into shells. Countercurrent handling of the various mixed acids (after fortification) increases the yield of the TNT and greatly decreases cost. Both the HNO_3 and $NO_2 \cdot HSO_3$ of the mono-spent acid are recovered as oxides of nitrogen by passing the spent acid down a tower against a rising column of steam The oxides are oxidized with air and dissolved in water to produce recycle nitric acid. The sulfuric acid in the bottom of the tower is safely concentrated for further use by countercurrent direct contact with hot combustion gas, thus harmlessly burning the nitrobodies.

The Schmid-Meissner and the Biazi processes are continuous-flow arrangements employing stirred-tank reactors that have been successfully adapted to the nitration of benzene and toluene in Europe. Each reactor in these processes is maintained at the optimum temperature, depending on the conversion and acid concentration. Advantages claimed for the processes include smaller and cheaper equipment for the same production, less hazardous operation since the quantities of TNT actually in the system at a given time are small, and better control of operating variables to minimize side reactions.

TNT acids, composition %

Compound	Mono reaction			Bi reaction		Tri reaction	
	Mixed acid	Spent acid	Cycle acid	Mixed acid	Spent acid	Mixed acid	Spent acid
H_2SO_4	50.4	56.5	54	53.9		40	64.5
HNO_3	14.5	3.6	6.6	13.8	2.5	60	3.9
$NO_2 \cdot HSO_4$	12.8	14.5	13.3	13			15
Nitrobenzene	2.5	0.6	2.3	11.3	Reduced		14.9
H_2O	19.8	25.0	26.6	8			1.9

Tetryl, ammonium picrate, and cyclonite (or RDX) are also widely employed as explosives. Cyclonite is cyclotrimethylenetrinitramine, which is relatively safe and a stronger explosive than TNT The latest process used in the United States involves continuous nitration by passing hexamethylenetetramine through a 4-in Pyrex glass pipe approximately 600 ft long and introducing strong nitric acid, ammonium nitrate, and acetic anhydride at appropriate intervals. Cyclonite yields are 70% and higher with strong acetic acid as a by-product.

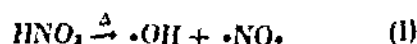


$$C_6H_{12}N_4 + 4HNO_3 + 2NH_4NO_3 + 6(CH_3CO)_2O \rightarrow 2C_3H_6O_6N_6 + 12CH_3COOH$$

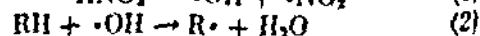
Liquid-phase nitration of paraffins. The liquid-phase nitration of paraffins is a relatively recent development employed for the production of nitrocyclohexane and 2,2-dinitropropane. Nitric acid reacts with decreasing reactivity at tertiary, secondary, and primary hydrogen atoms. The optimum operating variables are approximately as follows. temperature, 200°C; pressure, 1000 psi; and 70% nitric acid. The reaction mechanism may be both ionic and free radical in nature.

Vapor-phase nitration of paraffins. In 1955, the first commercial plant for producing nitroparaffins using a continuous-flow process was completed Propane is nitrated with nitric acid to produce 1-nitropropane, 2-nitropropane, nitroethane, and nitromethane These compounds find many uses as solvents, intermediates, and fuels

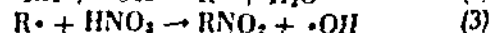
All lighter paraffins can be nitrated in the vapor phase, but ethane and especially methane nitrate only with difficulty. Propane and higher paraffins that contain secondary or tertiary hydrogens nitrate readily at 0–100 psi and at approximately 375–440°C in 0.5–2.0 sec when an excess of the alkane is employed Under these conditions, any hydrogen or alkyl group of the alkane can be replaced with a nitro group, and conversions of nitric acids to nitroparaffins are as high as 40%. The reaction occurs by a free-radical and probably a chain mechanism as follows:



$$HNO_3 \xrightarrow{\Delta} \cdot OH + \cdot NO_2 \qquad (1)$$



$$RH + \cdot OH \rightarrow R\cdot + H_2O \qquad (2)$$



$$R\cdot + HNO_3 \rightarrow RNO_2 + \cdot OH \qquad (3)$$

where reaction (1) is the chain-initiating step, and reactions (2) and (3) are the chain steps

Conversions as high as 70% are possible when oxygen or halogens are added in relatively small amounts to the reactant gases in order to increase the free-radical concentrations of the reacting gases. The nitroparaffin product obtained contains a higher concentration of nitromethane when oxygen is used and when the temperature in the reactor is relatively nonisothermal, but more nitropropanes are produced when halogens are employed and with good temperature control. In all cases, but especially when oxygen is employed, some oxygenated hydrocarbons are produced, but no disubstitute

nitro compounds are obtained with the vapor-phase process. Difficulty has been encountered in maintaining temperature control in a tubular reactor, but the exothermic heat of reaction can be used to furnish the latent heat for vaporizing cold liquid nitric acid which is sprayed into the hot gases. This reactor, although essentially adiabatic, maintains good temperature control.

Miscellaneous nitrations. Nitrogen dioxide can be used to nitrate paraffins and olefins, but apparently no commercial processes are employed. Propane reacts with nitrogen dioxide to produce a nitroparaffin product with a high concentration (about 72%) of 2-nitropropane, but conversions are low. Nitrogen dioxide undergoes addition reactions with olefins to produce dinitro alkanes and nitro nitrates. Olefins can be nitrated with nitric acid to produce nitroolefins, nitro alcohols, and the subsequent nitro nitrate esters.

The classic Victor Meyer method involving the reaction of silver nitrite with alkyl chlorides is used for the laboratory preparation of aliphatic nitro compounds. A technique employing alkali metal nitrites such as $NaNO_2$, $LiNO_2$, and KNO_2 has been perfected, in which a solvent such as ethylene glycol, water, or other hydrogen-donating solvent is used. Other nitrating agents that find limited

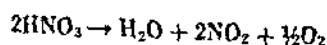
...echnology, 15 vols., 1947–1956.

Nitric acid

A strong mineral acid having the formula HNO_3. Pure nitric acid is a colorless liquid with a density of 1.52 at 25°C; it freezes at −47°C. Nitric acid is used in the manufacture of fertilizers, plastics, dyes, and lacquers. It may be prepared by the action of concentrated sulfuric acid on a nitrate or by the interaction of nitrogen dioxide (NO_2) with water. The second reaction is the basis for two commercial processes for manufacturing the acid. In the arc process, nitric oxide (NO) is produced by passing air through an electrical discharge. Upon cooling in air the nitric oxide forms nitrogen dioxide.

In the Ostwald process, ammonia (NH_3) produced from atmospheric nitrogen is oxidized to nitrogen dioxide.

Nitric acid is unstable. It decomposes upon heating and exposure to light, to water, nitrogen dioxide, and oxygen.



$$2HNO_3 \rightarrow H_2O + 2NO_2 + \tfrac{1}{2}O_2$$

It is a strong oxidizing agent, oxidizing carbon to carbon dioxide, sulfur to sulfuric acid, and phosphorus to phosphoric acid. It reacts with most metals, the products depending upon the position of the metal in the electromotive series. *See* AMMONIA; NITROGEN; NITROGEN OXIDES; OXIDIZING AGENT.

[F.J.]

Nitride

A binary compound of nitrogen with elements less electronegative than nitrogen. Common practice, however, is to exclude azides (such as NaN_3) in which specialized bonding exists among the nitrogen atoms, and the binary compounds with hydrogen, the halogens, and the oxygen group elements. With this limitation, essentially solid nitrides will be encountered.

Direct reaction of group Ia metals with nitrogen yields azides, which can be decomposed by heating (cautiously, to avoid explosion) to form the nitrides Li_3N, Na_3N, K_3N, Rb_3N, and Cs_3N. These nitrides decompose to nitrogen gas and the elements in the vicinity of 400°C; they react with water vapor to liberate ammonia (NH_3) and form the metal hydroxides.

The nitrides of the group IIa elements (Be_3N_2, Mg_3N_2, Ca_3N_2, Sr_3N_2, and Ba_3N_2) may be prepared by direct reaction of the element with nitrogen



$$N_2 + 3Be \rightarrow Be_3N_2$$

or with ammonia



$$2NH_3 + 3Be \rightarrow Be_3N_2 + 3H_2$$

at elevated temperatures. A convenient preparation is by heating the amides, for example,



$$3Ba(NH_2)_2 \rightarrow Ba_3N_2 + 4NH_3$$

These nitrides begin to melt and decompose to the elements and nitrogen gas in the vicinity of 1000°C, except for Be_3N_2 which melts at about 2200°C; they react with water in the same fashion as noted above for the alkali metal nitrides.

The nitrides of the group IIIa metals and of the lanthanide and actinide elements (ScN, YN, LaN, CeN, PrN, NdN, GdN, ThN, Th_2N_3, PaN_2, UN, U_2N_3, UN_2, NpN, and PuN) are quite stable to temperatures in the vicinity of 1500°C and much higher in some cases. The nitrides of the group IVa metals (TiN, ZrN, ... roup ... $_2$N) ... VIa ... Mo_2N, and W_2N) and of group VIIa (Mn_4N, Mn_5N_2, and Mn_3N_2) are less stable. All of these may be prepared by direct reaction of the metal with nitrogen gas. The compositions of a number of these nitrides vary over an appreciable range from the simple definite proportions indicated by the form...

... are of very low stability and are best prepared by reaction of the metal with ammonia gas. Ni_3N will decompose to the metal and nitrogen gas at 1 atm pressure at about 450°C. Cu_3N, Zn_3N_2, Cd_3N_2 (indirect preparation), and InN are even less stable. GaN, however, is quite stable.

The elements Ag, Au, Hg, Tl, Sn, Pb, Sb, and Bi do not form nitrides.

and yields than does nitrobenzene because of partial oxidation.

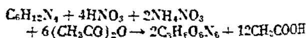
Trinitrotoluene (*TNT*) is a powerful explosive which is usually produced in a three-step nitration in order to secure economy of the nitrating acids, to reduce oxidation, and to secure best yields with only two nitrators necessary. The first nitration is carried out in the so-called mono house at 140°F, introducing toluene under agitation and with cooling into the cycle acid, followed by mono-mixed acid. The use of the cycle acid (last of the spent acid from previous batch) saves some nitric acid and increases the initial volume to secure better agitation and cooling. The resulting nitration product, so called *mono oil*, is conducted to the bi-tri house where binitration is performed at 90–180°F using bi-mixed acid. The bi-spent acid is settled, drawn off, and fortified to make mono-mixed acid. The binitration product is nitrated in the same kettle to TNT starting at 180° and ending at 230°F by the use of tri-mixed acid, following which the tri-spent acid is drawn off and sent to the fortifier for further use. The hot tri oil is withdrawn from the nitrator, crystallized in water, and washed in suspension with dilute soda ash solution followed by acidulated 16% Na_2SO_3 (Sellite) to remove undesirable isomers. After a water wash, the TNT is melted, washed, solidified or grained, and flaked or filled into shells. Countercurrent handling of the various mixed acids (after fortification) increases the yield of the TNT and greatly decreases cost. Both the HNO_3 and $NO_2 \cdot HSO_3$ of the mono-spent acid are recovered as oxides of nitrogen by passing the spent acid down a tower against a rising column of steam. The oxides are oxidized with air and dissolved in water to produce recycle nitric acid. The sulfuric acid in the bottom of the tower is safely concentrated for further use by countercurrent direct contact with hot combustion gas, thus harmlessly burning the nitrobodies

The Schmid-Meissner and the Biazi processes are continuous-flow arrangements employing stirred-tank reactors that have been successfully adapted to the nitration of benzene and toluene in Europe. Each reactor in these processes is maintained at the optimum temperature, depending on the conversion and acid concentration. Advantages claimed for the processes include smaller and cheaper equipment for the same production, less hazardous operation since the quantities of TNT actually in the system at a given time are small, and better control of operating variables to minimize side reactions.

TNT acids, composition %

Compound	Mono reaction			Bi reaction		Tri reaction	
	Mixed acid	Spent acid	Cycle acid	Mixed acid	Spent acid	Mixed acid	Spent acid
H_2SO_4	50.4	56.3	54	53.9		40	64.5
HNO_3	14.5	3.6	6.6	13.8	2.5	60	3.9
$NO_2 \cdot HSO_4$	12.8	14.5	13.3	13			15
Nitrobenzene	2.3	0.6	2.5	11.2	Reduced		14.9
H_2O	19.8	25.0	26.6	8			1.9

Tetryl, ammonium picrate, and cyclonite (or RDX) are also widely employed as explosives. Cyclonite is cyclotrimethylenetrinitramine, which is relatively safe and a stronger explosive than TNT. The latest process used in the United States involves continuous nitration by passing hexamethylenetetramine through a 4-in. Pyrex glass pipe approximately 600 ft long and introducing strong nitric acid, ammonium nitrate, and acetic anhydride at appropriate intervals. Cyclonite yields are 70% and higher with strong acetic acid as a by-product.

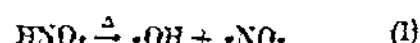


$$C_6H_{12}N_4 + 4HNO_3 + 2NH_4NO_3 + 6(CH_3CO)_2O \rightarrow 2C_3H_6O_6N_6 + 12CH_3COOH$$

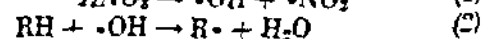
Liquid-phase nitration of paraffins. The liquid-phase nitration of paraffins is a relatively recent development employed for the production of nitrocyclohexane and 2,2-dinitropropane. Nitric acid reacts with decreasing reactivity at tertiary, secondary, and primary hydrogen atoms. The optimum operating variables are approximately as follows: temperature, 200°C; pressure, 1000 psi; and 70% nitric acid. The reaction mechanism may be both ionic and free radical in nature.

Vapor-phase nitration of paraffins. In 1955, the first commercial plant for producing nitroparaffins using a continuous-flow process was completed. Propane is nitrated with nitric acid to produce 1-nitropropane, 2-nitropropane, nitroethane, and nitromethane. These compounds find many uses as solvents, intermediates, and fuels.

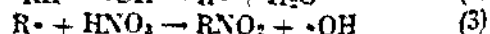
All lighter paraffins can be nitrated in the vapor phase, but ethane and especially methane nitrate only with difficulty. Propane and higher paraffins that contain secondary or tertiary hydrogens nitrate readily at 0–100 psi and at approximately 375–440°C in 0.5–2.0 sec when an excess of the alkane is employed. Under these conditions, any hydrogen or alkyl group of the alkane can be replaced with a nitro group, and conversions of nitric acids to nitroparaffins are as high as 40%. The reaction occurs by a free-radical and probably a chain mechanism as follows:



$$HNO_3 \xrightarrow{\Delta} \cdot OH + \cdot NO_2 \quad (1)$$



$$RH + \cdot OH \rightarrow R\cdot + H_2O \quad (2)$$



$$R\cdot + HNO_3 \rightarrow RNO_2 + \cdot OH \quad (3)$$

where reaction (1) is the chain-initiating step, and reactions (2) and (3) are the chain steps.

Conversions as high as 70% are possible when oxygen or halogens are added in relatively small amounts to the reactant gases in order to increase the free-radical concentrations of the reacting gases. The nitroparaffin product obtained contains a higher concentration of nitromethane when oxygen is used and when the temperature in the reactor is relatively nonisothermal, but more nitropropanes are produced when halogens are employed and with good temperature control. In all cases, but especially when oxygen is employed, some oxygenated hydrocarbons are produced, but no disubstitut

media it may be an azoxy, azo, or hydrazo derivative Dinitrobenzene may be partially reduced to m nitroaniline. Ortho and para isomers are formed by nitrating chlorobenzene, and subsequently replacing the Cl with NH_2 from ammonia. The use of water instead of ammonia gives 2,4,6-trinitrophenol or picric acid, a high explosive. Complete nitration of toluene gives trinitrotoluene (TNT).

Aliphatic nitro compounds are prepared with difficulty, and have grown in importance only since the development of vapor-phase nitration of hydrocarbons with nitric acid vapors at 420°C. Other preparative methods include the oxidation of oximes and reaction of alkyl halides with sodium nitrite The aliphatic nitro compounds, or nitroalkanes, are colorless, high boiling, and soluble in organic solvents but only slightly soluble in water. The electron attractive nitro groups decrease C—H bond strength on the adjacent (α-) carbon, making nitroalkanes somewhat acidic and prone to hydrogen displacement Nitroalkanes may be reduced to amines or hydrolyzed slowly to acids; unlike aromatic nitro compounds, they do not explode readily. *See* NITROPARAFFIN

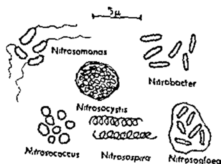
Nitroso compounds contain the —NO group attached to carbon or nitrogen Many are unstable intermediates, for example, nitrosobenzene formed during the red...

...ss and crystalline, but it forms a green solution. Tertiary amines will react with nitrous acid to form amine salts which lose water to give $R_3N—NO$ Nitroso compounds give identifying red, blue, and white colors with primary, secondary, and tertiary nitro compounds respectively. *See* NITRATION.

[A L.H.]

Nitrobacteriaceae

The nitrifying bacteria, a family of the order Pseudomonadales which live autotrophically, derive energy from the oxidation of either ammonia to nitrite, or nitrite to nitrate, and obtain carbon for growth from carbon dioxide They are the agents of nitrate formation in nature These bacteria are classified in the order Pseudomonadales because their motile cells have polar flagella, and placed in the suborder Pseudomonadineae because they



Representative genera of the Nitrobacteriaceae (V. B D Skerman)

are not photosynthetic. Seven genera are recognized; six of these are shown in the illustration.

Five genera of ammonia-oxidizing bacteria have been described: *Nitrosomonas, Nitrosococcus, Nitrosocystis, Nitrosospira* and *Nitrosogloea*. Very little is known about the last four genera

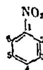
Nitrosomonas europaea Winogradsky is the best-known ammonia oxidizer and is widely distributed in arable soils. It has oval cells, about 1 by 0.8 micron (μ), which in some strains are massed together in a zoogloea, a colony of organisms embedded in a slimy substance. Motile "swarm cells" with one polar flagellum sometimes appear. This species does not form spores but grows in characteristic tiny colonies, only 100μ in diameter. *Nitrosomonas monocella* Nelson does not form a zoogloea, and has motile cells, each with a very long flagellum

Two nitrite-oxidizing genera are known, *Nitrobacter* and *Nitrocystis*. The common soil species is *Nitrobacter winogradskyi* Winslow et al., which is named for the Russian bacteriologist, Sergei Winogradsky (1856–1953), who first isolated nitrifying bacteria and discovered their autotrophic nature. It has oval cells about 0.8 by 0.7 μ, does not form a zoogloea, is nonsporeforming and nonmotile. The colonies are about 200μ in diameter. There is also a motile species, *Nitrobacter agile* Nelson. *See* BACTERIA, TAXONOMY OF; BACTERIAL NUTRITION; NITROGEN CYCLE; PSEUDOMONADALES; PSEUDOMONADINEAE, SCHIZOMYCETES; SOIL MICROBIOLOGY.

[J M.]

Nitrobenzene

A very pale yellow liquid with a sweet, but sickening odor. It boils at 210.9°C and freezes at 5.6 to



5.7°C. It is produced by the nitration of benzene with a mixture of nitric and sulfuric acids.

Nitrobenzene undergoes substitution reactions but requires more vigorous conditions than does benzene Substitution takes place at the meta (3) position.

Most of the nitrobenzene produced is reduced to aniline, but other dye intermediates including benzidine and metanilic acid are also prepared from it. The great toxicity of nitrobenzene impairs its usefulness as a solvent for organic compounds. If it is absorbed through the skin and if the vapor is inhaled, it may produce cyanosis. *See* AROMATIC HYDROCARBON, BENZENE; NITRATION.

[C K.B]

Nitrofuran

One of a group of 5-nitro substituted compounds of furan which have antimicrobial properties Since 1925 furan and its derivatives have been investigated for possible antimicrobial properties. In 1944 it was shown that a 5-nitro substituent greatly

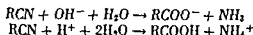
AlN, Si_3N_4, Ge_3N_4, and P_3N_5 are quite stable, and BN exceptionally so (sublimation point above 3000°C). *See* AZIDE; NITROGEN; OXIDE. [R.K.E.]

Nitrile

One of a group of organic chemical compounds, of general formula R—C≡N. The aliphatic nitriles containing up to 14 carbon atoms are liquids of high dielectric constant, and they are used as solvents.

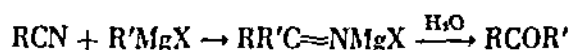
A nitrile is named from the acid to which it can be hydrolyzed by adding the suffix onitrile to the acid stem or as a cyanide of the group attached to CN. Thus, CH_3—CN is acetonitrile or methyl cyanide.

Nitriles hydrolyze to acids in either basic or acidic solution. They may be reduced by chemical agents or by catalytic hydrogenation to primary amines (often accompanied by secondary amines in the second case). Nickel and cobalt are the best catalysts:



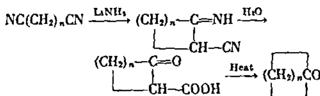
$$RCN + OH^- + H_2O \rightarrow RCOO^- + NH_3$$
$$RCN + H^+ + 2H_2O \rightarrow RCOOH + NH_4^+$$

Grignard reagents add to nitriles to give ketones (after hydrolysis):

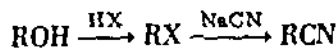


$$RCN + R'MgX \rightarrow RR'C{=}NMgX \xrightarrow{H_2O} RCOR'$$

In the presence of lithium amides α,ω-dinitriles undergo base-catalyzed condensation to cyclic cyanoimines, which can be hydrolyzed and decarboxylated to cyclic ketones (Thorpe reaction). Cyclic ketones with very large rings can be prepared by this method



The formation of cyanides from alkyl halides is an important chain-lengthening reaction in organic synthesis. It proceeds frequently from an available alcohol:



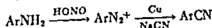
$$ROH \xrightarrow{HX} RX \xrightarrow{NaCN} RCN$$

The reaction is practical only with primary aliphatic halides, since the alkali cyanides are fairly strong bases and eliminate HX from secondary or tertiary alkyl halides.

Aromatic nitriles are made by replacement of a sulfonate group with alkali cyanide and by displacement of a diazotized primary amine group with the cyanide group in the presence of copper(I) cyanide (Sandmeyer reaction) or copper powder (Gattermann reaction).



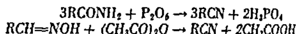
$$ArSO_3Na + NaCN \xrightarrow{Fusion} ArCN + Na_2SO_3$$



$$ArNH_2 \xrightarrow{HONO} ArN_2^+ \xrightarrow[NaCN]{Cu} ArCN$$

Gattermann reaction

The dehydration of acid amides or oximes with phosphorus pentoxide or acetic anhydride in either the aliphatic or aromatic series serves as another preparative method.



$$3RCONH_2 + P_2O_5 \rightarrow 3RCN + 2H_3PO_4$$
$$RCH{=}NOH + (CH_3CO)_2O \rightarrow RCN + 2CH_3COOH$$

For properties of two important members of the group, *see* ACRYLONITRILE; ADIPONITRILE. *See also* AMINE; CARBOXYLIC ACID; OXIME. [L.B.C.]

Nitrite

The negative ion, NO_2^-, derived from the unstable nitrous acid, HNO_2. Because of the intermediate oxidation state of nitrogen in nitrites (3+), the ion can act as either an oxidizing or reducing agent

Most nitrites are water-soluble. Sodium and potassium nitrites are quite stable and find extensive use in dyestuff manufacture and organic synthesis.

Sodium nitrite is produced by passing nitric oxide and nitrogen dioxide into sodium hydroxide as follows:



$$NO + NO_2 + 2NaOH \rightleftharpoons 2NaNO_2 + H_2O$$

When nitrite solutions are acidified by strong acids, this reaction is reversed with the evolution of a mixture of gases, and some nitrate may also be formed in the decomposition.

The presence of the nitrite ion is detected by the formation of the brown $Fe(NO)^{++}$ ion. *See* NITRATE; NITROGEN. [E.E.W.R.]

Nitro and nitroso compounds

Nitro compounds are derivatives of organic hydrocarbons having one or more $—NO_2$ groups with nitrogen to carbon bonding They differ from the oxygen-linked nitrites which are esters The group lacks enough electrons to form double bonds with both oxygens. However, both oxygens react alike, hence the bond is regarded as a resonance hybrid of single and double bonds

Aromatic nitro compounds, known for over 100 years, have been used chiefly as dye intermediates, explosives, and pharmaceuticals They are formed readily by the reaction of aromatic compounds with nitric acid; H is replaced by the $—NO_2$ group, for example, $—NO_2$ Nitro groups are electron-attracting, they impede the introduction of further substituents, and direct them to meta positions Nitration is aided by elevated temperature and by the presence of concentrated sulfuric acid, because of the latter's affinity for the water formed as well as its reaction with nitric acid to form nitronium ion (NO_2^+). This ion readily displaces an aromatic H atom. Compounds having electron-donating groups (CH_3— in toluene) are nitrated readily in the ortho and para positions *See* NITROBENZENE

Nitro compounds may be reduced by hydrogen to form primary amines. In cool neutral solutions the product may be hydroxylamine, and in a

pounds. The principal methods for doing this are the Haber process for the direct synthesis of ammonia from nitrogen and hydrogen, the electric-arc process, which involves the direct combination of N_2 and O_2 to nitric oxide, and the cyanamide process. Nitrogen is also used for filling bulbs of incandescent lamps and, in general, wherever a relatively inert atmosphere is required.

Atomic properties. The outer electron shells of the atoms of group Va elements have configurations of the type ns^2np^3. The normal configuration of the nitrogen atom is $1s^22s^22p^3$. Nitrogen is the most electronegative of the elements of this family (3.0 on the Pauling scale) and is a typical nonmetal in its reactions. The ionization potentials of the nitrogen atom are sufficiently high (first, 14.54 ev; second, 29.605; third, 47.426; fourth, 77.450; fifth, 97.863; sixth, 551.925) to prevent the nitrogen atom from forming positive ions under the ordinary conditions of chemical reaction. However, nitrogen atoms may, under some conditions, take up electrons to form N^{3-} ions. The crystal radius of this ion (Pauling scale) is 1.71 A. The covalent radius of tricovalent nitrogen on the Pauling scale is 0.74 A.

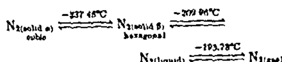
Nuclear properties. Nitrogen, as it occurs in nature, consists of two isotopes, N^{14} and N^{15}, in the abundance ratio of 99.635 to 0.365. In addition, the radioactive isotopes N^{12}, N^{13}, N^{16}, and N^{17} have been made by a variety of nuclear reactions. The first two are positron emitters and have half-lives of 0.0125 sec and 9.93 min, respectively. N^{16} and N^{17} are electron emitters and have half-lives of 7.35 sec and 4.14 sec, respectively. Unfortunately, none of these has a sufficiently long half-life for convenient use as a tracer. N^{15}, however, has been employed as a tracer by using nitrogen in which N^{15} has been concentrated and by following the reaction by mass-spectrometric techniques.

Molecular nitrogen. At standard temperature and pressure, elemental nitrogen exists as a gas with a density of 1.25046 g/l. This value indicates that the molecular formula is N_2. The N_2 molecule in its ground state has a magnetic susceptibility of -0.430×10^{-6} at 25°C, and therefore, has no resultant electronic angular momentum of either the orbital or spin variety. The electronic formula :N⋮N:, indicating a triple covalent bond, is commonly written for the molecule. The interatomic forces in the N_2 molecule are very high, as is indicated by a comparison of the interatomic distance in N_2, 1.095 A, with twice the single-bond covalent radius of nitrogen (2 × 0.74 A = 1.48 A). The energy of dissociation of the N_2 molecule into atoms as determined spectroscopically is 170.275 kcal/mole at 0°K. Spectroscopic studies indicate that, at ordinary temperatures, molecular nitrogen consists of molecules with symmetrical and antisymmetrical nuclear spins in the ratio 2:1, respectively. Because the nitrogen molecule is both very stable and highly symmetrical, intermolecular forces are very small. The following phase changes have been determined for molecular

Table 1. Properties of nitrogen

Property	Value
Heat of transformation (α–β)	54.71 cal/mole
Heat of fusion	172.3 cal/mole
Heat of vaporization	1332.9 cal/mole
Critical temperature	126.26 ± 0.04°K
Critical pressure	33.54 ± 0.02 atm
Density: α-form	1.0263 g/ml at −252.6°C
β-form	0.8792 g/ml at −210.0°C
Liquid	1.1607 − 0.00457 (T = abs. temp.)

nitrogen at standard atmospheric pressure (Table 1).

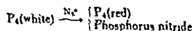


$$N_{2(\text{solid }\alpha)\,\text{cubic}} \xrightleftharpoons{-237.45°C} N_{2(\text{solid }\beta)\,\text{hexagonal}} \xrightleftharpoons{-209.96°C} N_{2(\text{liquid})} \xrightleftharpoons{-195.78°C} N_{2(\text{gas})}$$

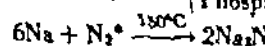
Elemental nitrogen is quite unreactive toward most common substances at ordinary temperatures. At high temper[illegible]

moderately high temperatures and pressures in the presence of a catal[illegible]

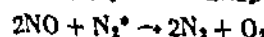
high-frequency electrodeless discharge, it is partially changed to an activated, unstable condition, from which, on standing, it returns to its normal state with the emission of a golden-yellow afterglow. Activated nitrogen is more reactive than ordinary nitrogen, as indicated by the following reactions:



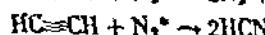
$$P_4(\text{white}) \xrightarrow{N_2^*} \begin{cases} P_4(\text{red}) \\ \text{Phosphorus nitride} \end{cases}$$



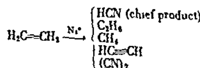
$$6Na + N_2^* \xrightarrow{150°C} 2Na_3N$$



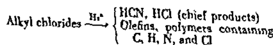
$$2NO + N_2^* \rightarrow 2N_2 + O_2$$



$$HC{\equiv}CH + N_2^* \rightarrow 2HCN$$



$$H_2C{=}CH_2 \xrightarrow{N_2^*} \begin{cases} \text{HCN (chief product)} \\ C_2H_6 \\ CH_4 \\ HC{\equiv}CH \\ (CN)_2 \end{cases}$$

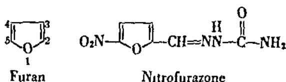


$$\text{Alkyl chlorides} \xrightarrow{N_2^*} \begin{cases} \text{HCN, HCl (chief products)} \\ \text{Olefins, polymers containing} \\ \text{C, H, N, and Cl} \end{cases}$$

Active nitrogen is believed to consist principally of nitrogen atoms in the ground state, and its superior chemical reactivity is believed to result from the presence of reactive free nitrogen atoms.

Compounds of nitrogen. The elements of the nitrogen family exhibit in their compounds three principal oxidation states, −3, +3, and +5. These

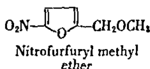
enhances the potency in a series of furans, and the most active compound was 5-nitro-2-furaldehyde semicarbazone, later called nitrofurazone.



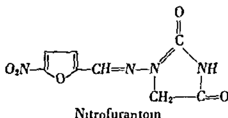
Furan Nitrofurazone

Nitrofurazone has a broad spectrum of antibacterial potency. A concentration of one part in 50,000 has a bactericidal action effective against many types of bacteria. Because of its systemic toxicity, it is used mainly as a topical germicide. Its only shortcoming in this application is its tendency to produce allergic manifestations in certain sensitized individuals

Two other clinically important nitrofuran derivatives have been introduced. One is nitrofurfuryl methyl ether, used as a topical fungicide The other is the first nitrofuran designed for systemic administration, *N*-(5-nitro-2-furfurylidene)-1-aminohydantoin, or nitrofurantoin.



Nitrofurfuryl methyl ether



Nitrofurantoin

This compound is advocated particularly for refractory urinary tract infections because, in addition to its broad antibacterial spectrum, it concentrates in an acidic urine. *See* CHEMOTHERAPY.

[N J.C]

Nitrogen

Chemical element number 7, nitrogen, N, is the lightest element of group Va (nitrogen family) which includes nitrogen, phosphorus, arsenic, anti- [illegible]

[illegible] s the principal [illegible] by volume of [illegible] dry air), in which its concentration is a result of the balance between the fixation of atmospheric nitrogen by bacterial, electrical (lightning), and chemical (industrial) action, and its liberation through the decomposition of organic materials by bacteria or combustion. In the combined state, nitrogen occurs in a variety of forms. It is a constituent of all proteins (both plant and animal) as well as of many other organic materials Its chief mineral source is sodium nitrate. An important source of this mineral is located in the arid regions of northern Chile.

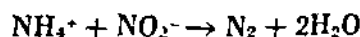
Preparation of the element. The methods for the preparation of elementary nitrogen may be grouped into two classes, separation from the atmosphere and decomposition of nitrogen compounds. The industrial method for the production of nitrogen is the fractional distillation of liquid air. Nitrogen containing about 1% argon and traces



of other inert gases may be obtained by the chemical removal of oxygen, carbon dioxide, and water vapor from the atmosphere by appropriate chemical reagents.

The following chemical reactions have been used to prepare nitrogen

When a saturated solution of sodium nitrite is mixed with a hot, saturated solution of ammonium chloride, the reaction which occurs is



$$NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$$

Ammonia gas is oxidized by passing it through bromine water, and the resulting gaseous mixture separated by passing it through a series of reagents to absorb unreacted bromine, water vapor, and ammonia The reaction is



$$2NH_3 + 3Br_2 \rightarrow N_2 + 6H^+ + 6Br^-$$

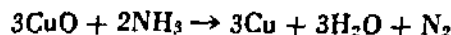
Other oxidants which may be used on ammonium salts include dichromate ion, ozone, fluorine, and manganese dioxide

The thermal decomposition of very dry barium azide or sodium azide yields spectroscopically pure nitrogen.



$$Ba(N_3)_2 \rightarrow Ba + 3N_2$$

Ammonia gas will react with hot metal oxides, for example,

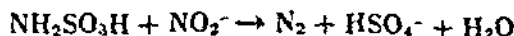


$$3CuO + 2NH_3 \rightarrow 3Cu + 3H_2O + N_2$$

to yield nitrogen.

Catalytic decomposition of ammonia on hot platinum produces nitrogen and hydrogen.

The reaction of sulfamic acid (or urea) with nitrite ion yields nitrogen.



$$NH_2SO_3H + NO_2^- \rightarrow N_2 + HSO_4^- + H_2O$$

Industrial application. Because of the importance of nitrogen compounds in agriculture and chemical industry, much of the industrial interest in elementary nitrogen has been in processes for converting elemental nitrogen into nitrogen c[illegible]

organic nitrogen is ammonified, however; a certain portion is retained in slowly decomposable complexes and becomes an integral part of the residual soil organic matter or humus

Nitrification. *Nitrification* is the bacterial oxidation of ammonia to nitrate, the chief source of readily available nitrogen for higher plants. It consists of two steps: first, ammonia is oxidized to nitrite by organisms of the genera *Nitrosomonas* and *Nitrosococcus*, and second, the resulting nitrite is oxidized to nitrate by *Nitrobacter* (*see* NITROBACTERACEAE). These highly specialized, autotrophic bacteria obtain their energy from these oxidations and their carbon from the carbon dioxide of the atmosphere. Generally the process of nitrite oxidation is faster than that of nitrite production, so that the level of nitrite in soil is too low to induce toxic effects. The few types of microbes known to be involved in nitrification require a much more restricted set of conditions for optimal activity than do the many types engaged in ammonification. A well-aerated, fertile, neutral to slightly alkaline soil will provide optimum conditions for nitrification. The nitrate so formed is utilized by plants and microorganisms, or may be lost from the soil by leaching. Under anaerobic conditions, nitrate may be reduced by the soil microflora.

Denitrification. In denitrification nitrate-nitrogen is reduced to nitrite, nitrous oxide, ammonia and, principally, molecular nitrogen. Under conditions of low oxygen tension a variety of soil microorganisms utilize nitrate as a source of oxygen and reduce it to forms which may be lost by leaching or may escape as gas into the atmosphere. The absence of oxygen as in waterlogged soil, and the presence of an abundant supply of soluble organic matter provide favorable conditions for this process. Normal agricultural soil is well-aerated, not too moist, and contains moderate amounts of organic matter or nitrate. Here denitrification is of little economic importance.

Nitrogen fixation. Molecular atmospheric nitrogen is returned to the soil primarily by man,

bacteria living in the nodules of leguminous plants and belong to the genus *Rhizobium*. The nonsymbiotic organisms are free-living bacteria which function either aerobically, as *Azotobacter*, or anaerobically, as *Clostridium* (*see* AZOTOBACTERACEAE, BACILLACEAE; RHIZOBIACEAE). Certain blue green algae such as *Nostoc*, *Anabaena*, and *Calothrix* species can fix atmospheric nitrogen. Several other groups of bacteria, including photosynthetic types, and fungi possess this characteristic to a limited degree, as has been demonstrated with isotopic nitrogen. The most important of the nitrogen fixing bacteria are those that produce nodules on the roots of legumes such as peas, beans, alfalfa, and clover. In this mutualistic association they may add well over 100 lb of atmospheric nitro-



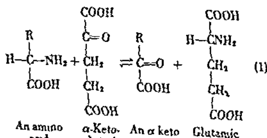
Fig. 2. Effect of inoculation with two strains of *Rhizobium leguminosarum* on growth of peas.

inoculated with active preparations of these organisms wherever legumes are grown. Figure 2 shows the effect of inoculating cultures of symbiotic nitrogen fixing bacteria on the development of peas. The nonsymbiotic forms, such as *Azotobacter* and *Clostridium*, fix much less nitrogen, with a fair average in fertile soils being about 10 lb/acre annually. Under certain conditions in the tropics, blue-green algae contribute significant amounts to the soil. *See* SOIL MICROBIOLOGY, SOIL MICROORGANISMS; SOIL MINERAL (MICROBIAL UTILIZATION). [H.K.]

Nitrogen excretion

In the quest for sufficient food energy to meet caloric requirements, animals ingest more nitrogen, largely as amino acids, than they require. Accordingly, the excess nitrogen ingested must be excreted in some form.

α-ketoglutaric acid to form glutamic acid. Thus



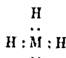
$$\mathrm{H{-}C(R){-}NH_2(COOH)} + \mathrm{COOH{-}C{=}O{-}CH_2{-}CH_2{-}COOH} \rightleftharpoons \mathrm{R{-}C{=}O(COOH)} + \mathrm{COOH{-}CH(NH_2){-}CH_2{-}CH_2{-}COOH} \quad (1)$$

An amino acid   α-Keto-   An α keto   Glutamic

Table 2. Compounds of nitrogen

Oxidation state	Examples
+5	N_2O_5, HNO_3, nitrates, NO_2X
+4	$N_2O_4 \rightleftharpoons 2NO_2$
+3	N_2O_3, HNO_2, nitrites, NOX, NX_3
+2	NO, Na_2NO_2, nitrohydroxylamates
+1	N_2O, $H_2N_2O_2$, hyponitrites
0	N_2
−⅓	HN_3, azides
−1	NH_2OH, hydroxylammonium salts
−2	NH_2NH_2, hydrazinium salts, hydrazides
−3	NH_3, ammonium salts, amides, imides, nitrides

states are readily correlated with electronic configurations. For example, the +3 state involves the p^3 electrons. All the elements of the nitrogen family form hydrides of the formula



+3 oxides, +5 oxides, +3 halides (MX_3), and, except for nitrogen and bismuth, +5 halides (MX_5) Nitrogen is the most nonmetallic element in the nitrogen family, in addition, it stands apart from the other elements of the family in three other important ways. (1) The valence shell of the nitrogen atom is limited to four orbitals (one *s* and three *p*; hence, the maximum covalence of nitrogen is four. (2) Nitrogen is capable of forming double bonds involving its *p* orbitals, whereas other members of the family have much less tendency to do this. Therefore, double bonds are common in nitrogen chemistry. (3) The small size of the nitrogen atom limits its coordination number because of steric factors; thus, *o*-nitric acid, H_3NO_4, has not been prepared. Table 2 lists the principal classes of inorganic nitrogen compounds. Thus, in addition *to the typical oxidation states of the family* (−3, +3, and +5), nitrogen forms compounds with a variety of additional oxidation states *See* AMINE, AMMINE; AMMONIA, ATMOSPHERIC GASES, PRODUCTION OF; HYDRAZINE; NITRIC ACID; NITRIDE; NITROGEN OXIDES [H.H.S.]

Bibliography: M. C. Sneed et al. (eds.), *Comprehensive Inorganic Chemistry*, vol. 5, 1956

Nitrogen cycle

The continuous cyclic exchange between combined nitrogen in the soil and molecular nitrogen in the atmosphere. *It includes all the transformations* concerned in the mineralization of nitrogenous organic substances and in the loss or gain of nitrogen by the soil

Soil nitrogen occurs naturally in organic and inorganic forms as a result of plant, animal, and microbial growth. *Nitrogen is stored in soil* primarily in organic combinations not utilizable by higher plants, but made available as ammonia through the activities of soil microorganisms The ammonia may be used by both higher plants and microorganisms either directly or after oxidation to nitrate-nitrogen. *Both ammonia and nitrate may* be lost from soil by leaching or through microbial action. Soil gains nitrogen chiefly through the addition of fertilizers and through microbial fixation of atmospheric nitrogen. The nitrogen cycle comprises the processes of ammonification, nitrification, denitrification, and nitrogen fixation.

Some authorities further subdivide ammonification into proteolysis (protein degradation to amino acids) and ammonification, and nitrification into nitritation (formation of nitrite from ammonia) and nitratation (formation of nitrate from nitrite).

Ammonification. Ammonification refers to the release of nitrogen as ammonia from organic compounds in plant, animal, and microbial residues This is accomplished chiefly under aerobic conditions through the participation of bacteria, fungi *actinomycetes, and other microscopic forms of life* The first step in the process involves the hydrolytic cleavage of proteins, nucleic acids, and related compounds to amino acids and other simple nitrogenous substances. These are then broken down to ammonia. Uric acid and urea, the excretory products of animals, are rapidly mineralized The ammonia-nitrogen liberated through microbial action is in excess of the requirements of these organisms for growth Consequently when a substance that is high in nitrogen, such as protein, is added to a soil considerable ammonia is liberated, whereas a substance relatively low in nitrogen, such as straw yields comparatively little, if any, ammonia. Furthermore, if an excessive amount of nonnitrogenous carbonaceous material, for example, carbohydrate is added to a soil, much available nitrogen, such as nitrate or ammonia, will be used by the rapidly developing microbial population, thus decreasing the available supply for higher plants. Not al

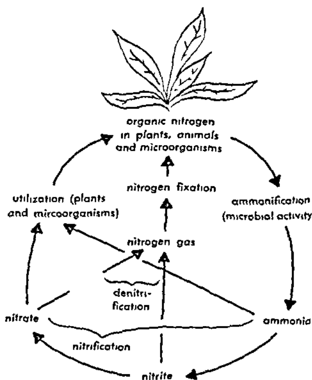


Fig. 1. Diagram of the nitrogen

$$\begin{array}{c}\\ COOH\\ |\\ C{=}O\\ |\\ CH_2\\ |\\ COOH\end{array} + \begin{array}{c}COOH\\ |\\ HCNH_2\\ |\\ CH_2\\ |\\ CH_2\\ |\\ COOH\end{array} \rightleftharpoons \begin{array}{c}\\ COOH\\ |\\ HCNH_2\\ |\\ CH_2\\ |\\ COOH\end{array} + \begin{array}{c}COOH\\ |\\ C{=}O\\ |\\ CH_2\\ |\\ CH_2\\ |\\ COOH\end{array} \tag{6}$$

Oxalo-acetic acid | Glutamic acid | Aspartic acid | α-Keto-glutaric acid

$$\begin{array}{lll} & & COOH\\ & & |\\ (CH_2)_3{-}NH & & CH_2\\ | & | & |\\ HCNH_2 & C{=}N{-} & CH\\ | & | & |\\ COOH & NH_2 & COOH\end{array} \rightarrow$$

Argininosuccinic acid

$$\begin{array}{ll}(CH_2)_3{-}NH & \\ | & |\\ HCNH_2 & C{=}NH\\ | & |\\ COOH & NH_2\end{array} + \begin{array}{c}HOOC \quad H\\ \diagdown\ \diagup\\ C\\ \|\\ C\\ \diagup\ \diagdown\\ H \quad COOH\end{array} \tag{7}$$

Arginine | Fumaric acid

$$\begin{array}{ll}(CH_2)_3{-}NH & \\ | & |\\ HCNH_2 & C{=}NH\\ | & |\\ COOH & NH_2\end{array} + H_2O \rightarrow \begin{array}{l}(CH_2)_3{-}NH_2\\ |\\ HCNH_2\\ |\\ COOH\end{array} + \begin{array}{c}NH_2\\ |\\ C{=}O\\ |\\ NH_2\end{array} \tag{8}$$

Arginine | Ornithine | Urea

The free energy change $\Delta F°$ required for the over-all reaction

$$CO_2 + 2\ \text{Ammonia} \rightarrow \text{Urea}$$

is −14,000 cal/mole As in many other biological systems, this energy is provided by utilization of adenosine triphosphate The cyclic nature of the system is illustrated in Fig 1 A second cycle is also operative wherein the 4-carbon dicarboxylic acid, fumaric acid, released in reaction (7) is re-converted to oxaloacetic so that it may be reutilized for aspartic acid formation and thus for another turn of the urea cycle See EXCRETION; URINARY SYSTEM [P.H]

Nitrogen fixation

A process to convert atmospheric nitrogen to chemical compounds that are useful as such or as fertilizers Arc processes have been important but now are used only where electricity is very cheap. For descriptions of more recent and important processes, see A...

In a ...

leaves ...

is cooled and scrubbed with water or alkali to ... nitric oxide. It ... cover the nitric oxide as nitric acid

In the Birkeland-Eyde process used in Norway, air passes through an alternating current (ac) arc flattened by a magnetic field. Furnaces use 3200–4000 kw, 5000 volts, 13,600 kwhr per ton HNO_3 at 80% power factor. Electrodes last 400-500 hours, and furnace linings, 4-6 months.

The Schoenherr-Hessberger process, also used in Norway, employs a very long ac arc around which air moves in a helical path. The arc is 7 m long in a 746-kw furnace. The power distribution is 3% to form NO, 30% to steam, 40% to hot water, 10% to cooling water, 17% as heat losses

The Pauling process used in France and the Tyrol uses a fan-shaped arc blown by the air. A 400-kw furnace at 4000 volts has an arc 1 m long See ELECTROCHEMICAL PROCESS; NITROGEN; NITROGEN CYCLE. [W.C.G.]

Bibliography: C. L. Mantell, *Industrial Electrochemistry*, 3d ed , 1950.

Nitrogen oxides

Chemical compounds of nitrogen and oxygen. Nitrogen and oxygen do not combine when mixed directly (as in air) but they do combine during chemical reactions of compounds containing them. A number of nitrogen oxides can be isolated which differ from one another ...

[illegible]

...ratory method for the preparation of each oxide, are given in Table 2

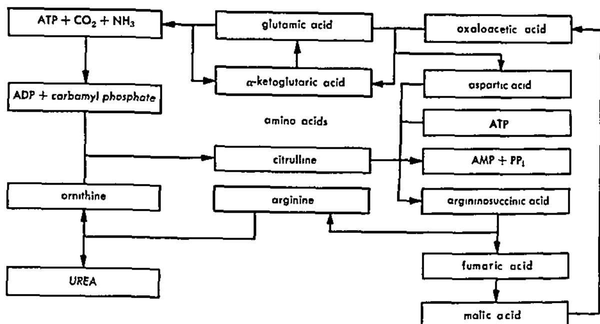
These structures show only the geometry of the molecules In most cases the N and O atoms are united by complex (double or triple) bonds.

The existence of three higher oxides has been postulated. They are nitrogen trioxide (NO_3) from reaction of ozone with dinitrogen tetroxide or pentoxide, dinitrogen hexoxide (N_2O_6) from reaction of fluorine with nitric acid, and an oxide NO_4 as an intermediate in the O^{18} isotope exchange between dinitrogen pentoxide and oxygen gases. The identity and properties of these three oxides are not fully established

Nitrous oxide and nitric oxide. When inhaled, nitrous oxide has anesthetic effects; in small

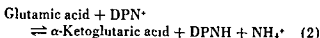
Table 1. Oxides of nitrogen and their properties

Name	Stoichiometric formula	Melting point, °C	Boiling point, °C
Nitrous oxide (dinitrogen monoxide)	N_2O	−90 8	−88 5
Nitric oxide (nitrogen monoxide)	NO	−163 6	−151 7
Dinitrogen trioxide	N_2O_3	−103	+3 5
Dinitrogen tetroxide (⇌ nitrogen dioxide)	N_2O_4 (⇌ NO_2)	−11 2	+21 2
Dinitrogen pentoxide	N_2O_5	+41	



Regenerative cycles in urea biosynthesis.

Under the influence of *glutamic dehydrogenase*, glutamic acid may be oxidized by the coenzyme diphosphopyridine nucleotide (DPN) with the re*formation of α-ketoglutaric acid plus* ammonia.



$$\text{Glutamic acid} + \text{DPN}^+ \rightleftharpoons \alpha\text{-Ketoglutaric acid} + \text{DPNH} + \text{NH}_4^+ \quad (2)$$

Many marine species simply excrete this ammonia as such in their urine, whereas most terrestrial animals first convert the ammonia into urea To accomplish this transformation, advantage is taken of the enzymically catalyzed metabolic sequence by which the amino acid arginine is synthesized from ornithine, a sequence common to almost all living forms. It is adapted for urea synthesis by mammals, reptiles, and other forms by the additional presence in liver of the enzyme arginase which catalyzes the hydrolysis of arginine to urea plus ornithine which is then available for recycling. *See* UREA.

Urea cycle. The initial step is catalysis of the formation of carbamyl phosphate from ammonia, CO_2, and adenosine triphosphate (ATP) by carbamyl phosphate synthetase. The mechanism of this reaction is not understood but it is known to require the presence of an *N*-acylglutamate and an additional molecule of ATP.

$$NH_3 + CO_2 + 2ATP \rightarrow H_2N\text{—}CO\text{—}OPO_3H_2 + H_2OPO_3^- + 2ADP \quad (3)$$

Carbamyl phosphate

The carbamyl phosphate is then caused to react with ornithine under the influence of ornithine transcarbamylase with the formation of citrulline plus orthophosphate, a reaction shown in Eq. (4). The second atom of nitrogen necessary for urea synthesis is then introduced in a complex reaction whereby aspartic acid and citrulline condense to form argininosuccinic acid, the energy being derived from another molecule of adenosine triphosphate with the formation of adenosine monophosphate and pyrophosphate, as is shown in Eq. (5).

$$\underset{\text{Ornithine}}{NH_2\text{—}(CH_2)_3\text{—}HCNH_2\text{—}COOH} + NH_2\text{—}COO\text{—}PO_3H_2 \rightleftharpoons \underset{\text{Citrulline}}{NH\text{—}CO\text{—}NH_2\text{—}(CH_2)_3\text{—}HCNH_2\text{—}COOH} + H_2PO_4^- \quad (4)$$

$$\underset{\text{Citrulline}}{(CH_2)_3\text{—}NH\text{—}C(=O)\text{—}NH_2;\ HCNH_2;\ COOH} + \underset{\text{Aspartic acid}}{COOH\text{—}HCNH_2\text{—}CH_2\text{—}COOH} \rightarrow \underset{\text{Argininosuccinic acid}}{(CH_2)_3\text{—}NH\text{—}C(NH_2){=}N\text{—}CH(COOH)\text{—}CH_2\text{—}COOH;\ HCNH_2;\ COOH} \quad (5)$$

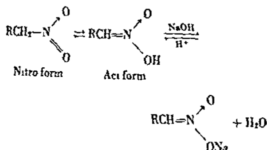
It is noteworthy that, like ammonia, the nitrogen of aspartic acid is derived from glutamic acid through the operation of a specific aspartic-glutamic transaminase as shown in Eq (6)

The product, argininosuccinic acid, is cleaved by an appropriate enzyme to form one mole each of arginine and fumaric acid as shown in Eq (7). With the hydrolysis of arginine to ornithine plus urea the cycle is completed and e ornithi available for the next round F

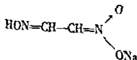
Nitroparaffin

One of a series of aliphatic compounds with relatively high boiling points and high molecular polarity. Such compounds are used as solvents for cellulose derivatives, waxes, and many substances of high molecular weight. Their chief use is as intermediates for the synthesis of other organic compounds.

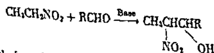
Direct nitration of propane at 400°C with nitric acid under 150 lb pressure furnishes the commercially available nitroparaffins: nitromethane, b.p. 101°C; [illegible]



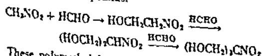
The presence of the aci (acid) form can account for the relatively low pH which is observed in aqueous suspensions of the fairly insoluble nitroparaffins. Tertiary nitroparaffins, R_3CNO_2, which lack an α hydrogen, do *not* display this behavior. Nitromethane reacts with strong bases in a special manner to give salts of methazonic acid.



Primary and secondary nitroparaffins undergo aldol type condensations with aldehydes to give nitro alcohols.

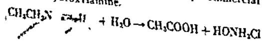


With formaldehyde, the condensation may be repeated as long as an α-hydrogen remains to yield polymethylol compounds.



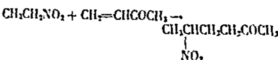
These polymethylol compounds can be reduced to the industrially important amino alcohols. *See* ALKANOLAMINE.

Acid hydrolysis of nitroparaffins yields the corresponding carboxylic acid and salts of hydroxylamine. This reaction provides a ready commercial route to hydroxylamine.



Halogenation of primary and secondary nitroparaffins proceeds readily in alkaline solution. This reaction is analogous to the halogenation of aldehydes and ketones. Only the α-hydrogens are replaced. The 1-chloro- and 1,1-dichloronitroparaffins find use as insecticides.

Primary and secondary nitroparaffins add to α,β-unsaturated systems, for example, in base-catalyzed Michael condensations.



See NITRATION. [L.B.C.]

Nobelium

Element 102, nobelium, No, is the heaviest element known to date. It was the latest of the synthetic elements to be made in the laboratory. The element decays by emitting an α particle, that is, a doubly charged helium ion. Approximately 100 atoms of the element have been produced.

The element is the tenth element heavier than uranium to be produced synthetically. It is the thirteenth member of the actinide series of elements, a rare-earthlike series of elements, which will be completed when element 103 has been produced.

For a discussion of the properties of the rare earths, *see* RARE-EARTH ELEMENTS.

The element was made by bombarding isotopes of curium, element 96, with carbon ions accelerated to energies of 100 million electron volts (Mev) in a heavy-ion linear accelerator or in a heavy-ion accelerating cyclotron. The curium atoms capture some of the high-energy carbon ions to form highly excited atoms of element 102. The excited atoms lose their excess energy by evaporating neutrons, and yield the final isotope of element 102 which is then isolated and identified.

Atoms having the same atomic number, and hence identical chemical behavior, but differing in atomic weight are isotopes of a given element. For a further discussion *see* ISOTOPE.

Since the isotopes of this element disintegrate rapidly after their formation, that is, they have a short half-life, special techniques were necessary to characterize the atoms of element 102 shortly after they were formed.

Table 2. Oxides of nitrogen, their formulas and preparation

Name	Structure	Preparation
N_2O	N—N—O	*Heat ammonium nitrate*
NO	N—O	Reduce nitric acid with copper
N_2O_3	$O-N-N(-O)-O$	Condense gaseous mixture of NO and NO_2
NO_2	$O-N-O$	Heat lead nitrate
N_2O_4	O_2N-NO_2	Heat lead nitrate
N_2O_5 *Gas*	$O_2N-O-NO_2$	Treat N_2O_4 with ozone
Solid	$NO_2^+\ NO_3^-$	

amounts it produces mild hysteria and hence is sometimes called laughing gas. It is colorless, is the least reactive of the oxides, and dissolves in water without chemical reaction Decomposition into nitrogen and oxygen occurs at an appreciable rate above 560°C.

The equilibrium



$$N_2 + O_2 \rightleftharpoons 2NO$$

lies entirely to the left at low temperatures. Some nitric oxide is formed in an electric arc, as in the technical production of nitric acid

With oxygen or air, nitric oxide is rapidly converted to nitrogen dioxide. Nitric oxide is colorless and is soluble in water without reaction. It is one of *the few "odd" molecules which contain an odd* number of electrons. Other such molecules (for example, nitrogen dioxide) readily form double molecules, but nitric oxide is exceptional. The gas is monomeric, although dimerization occurs in the *liquid, and solid* nitric oxide (which is blue) is almost entirely in the form of N_2O_2 molecules. As an odd molecule it has the ability to lose or gain one electron, thus giving the electrically charged ions NO^+ and NO^-. The important nitrosyl compounds contain these ions

Trioxide. Dinitrogen trioxide exists pure only in the solid state The dissociation



$$N_2O_3 \rightleftharpoons NO + NO_2$$

occurs partially in the blue liquid, and almost entirely in the vapor state at room temperature. It is the anhydride of nitrous acid; when the oxide (or an equimolecular mixture of NO and NO_2 gases) is dissolved in an alkaline solution, nitrite ion is produced.

Dioxide and tetroxide. The position of the equilibrium between nitrogen dioxide and dinitrogen tetroxide



$$N_2O_4 \rightleftharpoons 2NO_2$$

depends upon temperature and physical state The dioxide is a red-brown poisonous gas; the tetroxide is colorless. The colorless solid is entirely in the tetroxide state. In the liquid and gaseous states the tetroxide always contains some dioxide. Thus the liquid tetroxide is brown, although it contains less than 0.1% nitrogen dioxide The color of the gas becomes more intense with rising temperature; at 100°C the tetroxide is 90% dissociated into dioxide. At temperatures above 600°C further decomposition of nitrogen dioxide into nitric oxide occurs

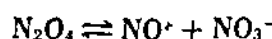


$$NO_2 \rightleftharpoons NO + \tfrac{1}{2}O_2$$

Dinitrogen tetroxide reacts readily with water to give an equimolecular mixture of nitrous and nitric acids. As temperature is raised the nitrous acid decomposes to nitric acid and nitric oxide These reactions are important in the technical production *of nitric acid by catalytic oxidation of ammonia* Dinitrogen tetroxide is an oxidizing agent comparable in strength to bromine, and is employed as such in the lead-chamber process for sulfuric acid In organic chemistry the tetroxide finds use as a special oxidizing agent (for example, in the production of sulfoxides and phosphine oxides) and as a nitrating agent.

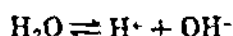
The tetroxide forms molecular addition compounds with many simple organic solvents, for example, esters, ethers, ketones, and nitriles.

Liquid dinitrogen tetroxide, alone or mixed with organic solvents, undergoes self-ionization



$$N_2O_4 \rightleftharpoons NO^+ + NO_3^-$$

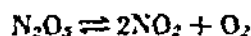
This is to be compared with the aqueous system



$$H_2O \rightleftharpoons H^+ + OH^-$$

For example, liquid dinitrogen tetroxide attacks some metals (alkali and alkaline-earth metals, zinc, cadmium, and mercury) to produce metal nitrate and evolve nitric oxide A scheme of reactions has been developed using the liquid tetroxide as a reaction medium with nitrosyl compounds as acids and nitrates as bases This medium is therefore valuable for the preparation of anhydrous metal nitrates and nitrato-coordination complexes

Pentoxide. The ionic nitronium nitrate structure $NO_2^+ \cdot NO_3^-$ found for N_2O_5 *in the solid state* accounts for its anomalously high melting point *In solution in sulfuric, nitric, or phosphoric acids the* oxide has the same ionic structure. Solid dinitrogen pentoxide readily volatilizes, and the molecular type of structure found in the gaseous state is observed also in solutions of the oxide in low dielectric solvents such as carbon tetrachloride and chloroform. Sodium metal reacts with the liquid oxide, liberating nitrogen dioxide and forming sodium nitrate. Gaseous dinitrogen pentoxide decomposes readily



$$N_2O_5 \rightleftharpoons 2NO_2 + O_2$$

and is a strong oxidizing agent. With water it is converted to nitric acid *See* NITROGEN; OXYGEN.

terms of the discrete variables used for single-frequency sounds.

The statistical description of random sound waves parallels that used for electrical noise (*see* NOISE, ELECTRICAL). The magnitude of the noise is usually specified in terms of its radiated intensity in a 1 cycle per second (cps) frequency band, also known as the intensity spectrum level. If the random noise has a relatively uniform distribution of intensity as a function of frequency, it is often described as intensity in a frequency band more than 1 cps wide. This may be done in terms of a constant bandwidth, such as 5, 50, or 500 cps, or in terms of a constant percentage bandwidth, such as 1/10, 1/3, or 1 octave of frequency. The most common usage in industrial noise control is specification of intensity in octave frequency bands. These bands usually cover the frequency range of 37–9600 cps, although the lower limit is often extended to 20 and the upper to 10,000 cps. Each of these latter bands is then more than one octave wide. The over-all intensity of a random noise is the sum of the intensities in all the frequency bands by which it is specified.

White noise. Random noise having the same intensity, in a 1-cps band, at every frequency in the range of interest is called white noise. Although white noise is a fairly common type of electrical noise, it is rarely encountered in acoustic noise. Most random acoustic noises tend to have a definite nonuniform distribution of intensity as a function of frequency, for example, *see* AIRCRAFT NOISE.

Ambient noise. The residual noise present at any location of interest is called ambient noise. It is the sum of all noises present. Thus ambient noise in an office could be the result of ventilating systems, distant conversations, office machines, and so forth. Background noise is a term often used to describe the ambient noise when a particular source of sound being studied is not in operation. *See* NOISE MEASUREMENT, UNDERWATER SOUND.

[W.J.C.]

Noise, electrical

Interfering and unwanted currents or voltages [illegible] an elect[illegible] ally [illegible] any [illegible] mit [illegible] present information. In such systems as telephone, radio, television, radar, radio-navigation, telemetering, electronic control, or electronic computing, the desired signals carrying intelligence may be masked or distorted by noise.

Noise may originate either externally to the device in which it appears, as atmospheric static, or internally, as thermal noise in a resistor. It may result either from natural phenomena, as do both the types of noise just mentioned, or from interference from man-made devices, such as nearby electric motors or generators.

Man-made interference can usually be nearly eliminated by good engineering design and proper location of equipment. Noise due to natural phenomena often cannot be reduced below certain fixed levels, and good engineering design can ensure only that the equipment will function as effectively as possible in the face of this irreducible noise. For example, a radio receiver cannot operate on received signals which are very weak (compared to some level determined by the thermal noise in the receiver), no matter how much amplification is used in the receiver, because the noise is amplified along with the signal.

Sources of noise. Noise may conveniently be classified as either random noise or nonrandom noise. Random noise is defined as noise which is not predictable, although it may exhibit statistical regularities.

Random noise in electron-tube circuits. Thermal noise is the random voltage which appears at the terminals of a resistor or any component with internal resistance because of the random motion of thermally excited electrons in the resistor. Shot noise is caused by fluctuations of current in a thermionic vacuum tube caused by random emission of electrons from its hot cathode. Partition noise is the result of fluctuations in current to one electrode in a multielement vacuum tube, caused by random division of the electron stream between two or more collecting electrodes (for example, the screen-grid and anode of a tetrode). Flicker noise is the result of low-frequency fluctuations in current in a vacuum tube apparently caused by relatively slowly changing emission conditions at various points on the cathode. Contact noise, the noise in carbon resistors and carbon microphones, for example, is caused by randomly varying fluctuations in resistance.

Random noise in semiconductors. Random noise also originates in transistors and other semiconductor devices. There are various mechanisms at work, and the terminology is not completely standard. Thermal noise, as above, is the noise caused by thermally excited, random motion of electrons. Shot noise, or generation recombination noise, is produced by fluctuations in free carrier densities when an electric field is applied. Excess noise or modulation noise is caused by slow fluctuations in conductivity. Noise also originates in photoconductors. Most mechanisms of random noise generation share the principle that observable gross currents and voltages are the result of many random actions at a microscopic level.

Radiated random noise. Any electric device which must receive electromagnetic radiation will pick up radiated random noise as well as signals. Electrical disturbances in the atmosphere cause noise which is very irregular in character, often appearing in sharp [illegible] TRICITY; ST. [illegible] ances, an an[illegible] noise which [illegible] origin, thermal radiation from the gases of the atmosphere, and thermal radiation from heavenly bodies and systems. This last is sometimes called interstellar noise (*see* RADIO ASTRONOMY). The Sun radiates noise at all

When a high-energy particle, such as a 100-Mev carbon ion, is captured by a curium atom, the resulting combined atom is caused to move in a forward direction by the original momentum of the carbon ion with enough velocity to separate it from the stationary curium atoms. Under appropriate conditions these recoil atoms of element 102 can be caught on adjacent catcher foils. By employing such a catcher technique, it is possible to separate the atoms of element 102 formed by the nuclear reactions from the curium target and identify them rapidly. In one case, in which the atoms of element 102 had a half-life of 3 sec, it was not possible to identify them directly. However, they decayed by α-particle emission to element 100. The process of α-decay gave enough energy to the resultant atoms of element 100 so that they in turn recoiled off the foil and were caught on another foil. By identifying the recoil atoms of element 100, it was possible to show that the atoms decaying with the 3-sec half-life were an isotope of element 102. *See* ACTINIDE ELEMENTS; NUCLEAR CHEMISTRY; NUCLEAR REACTION; RADIOACTIVITY; TRANSURANIUM ELEMENTS [P.R.F.]

Nocardiosis

An actinomycetous disease of man caused by several species of *Nocardia.* The lesions are purulent and granulomatous, usually involving the subcutaneous tissues, although a pulmonary form is also recognized. This latter form of nocardiosis is the most serious, often metastasizing to the viscera and central nervous system. *Nocardia asteroides,* the primary etiological agent, is easily found in the soil It is a gram-positive branching filament (1 μ in diameter), often forming granules in tissues Since many strains of *Nocardia asteroides* are also acid-fast, care must be taken not to confuse them with the tubercle bacillus *See* ACTINOMYCETALES

The organism grows readily on most laboratory media but requires special biochemical and pathogenicity studies for differentiation from saprophytic species of *Nocardia.* This disease may be treated with antibiotics and sulfonamides *See* ANTIBIOTIC; MYCOLOGY, MEDICAL, SULFA DRUGS. [L D H]

Noeggerathiales

An incompletely known and poorly defined group of vascular plants whose geologic range extends from Upper Carboniferous to Triassic. A rather heterogeneous assemblage of both foliar and vegetative organs, assignable to fairly well defined genera have been placed in the group, thus somewhat forcing the concept that these parts comprise a natural order of vascular plants. Internal anatomy is unknown, with exception of the possible Noeggerathialean [illegible] Noeggera-thialean genera i[illegible] *thiostrobus, Tingi*[illegible] The reproductive organs are strobiloid, with whorled organization, and vary from homosporous to heterosporous. In *Discinites* the number of megaspores may range from 1–16 per sporangium. Foliar organs vary from nonarticulate and fern like to anisophyllous four-ranked fronds The Noeggerathiales have been proposed in the evolutionary scheme in two remotely related groups of vascular plants, the Pteropsida and the Sphenopsida. Morphological evidence, however, owing to paucity of the fossil record, does not make it possible to place the group with certainty in any recognized major subdivision of the vascular plants. *See* PALEOBOTANY; PLANT KINGDOM. [E.S.B.]

Noise, acoustic

Unwanted sound. This definition of acoustic noise, while purely general, implies that some criterion must exist before the sound can be termed unwanted. Whether a sound is noise, insofar as humans are concerned, is a subjective matter.

Criteria. Considerable effort has been expended to analyze unwanted sounds in an effort to specify objective criteria for the subjective human reactions to noise These criteria have included annoyance, interference with speech, damage to hearing, and reduction in efficiency of work performance. For a discussion of these factors, *see* NOISE CONTROL; PSYCHOACOUSTICS.

Noise is usually thought of in terms of its effect on humans; an equally important aspect, however, is its effect on the fatigue or malfunction of physical structures and equipment. In these instances criteria can, in theory, be established on a completely objective basis. The "unwanted" aspect in the definition of noise applies here to the fact that it is generally considered undesirable to have a structure such as an aircraft experience fatigue, or that it is undesirable to have electronic equipment guiding the aircraft fail because of malfunctions brought on by intense sound waves.

A third major criterion for describing sound as noise arises in conjunction with the perception or detection of a wanted sound in the presence of other sounds which tend to mask it Thus in sonar the reflected sound from an object being detected is a signal which is wanted, whereas all other sound detected by the system is termed noise. *See* SONAR.

Physical specifications. The generality of the preceding definitions gives no clue to the physical specifications of sound waves called noise. The sound can be composed of definite pure-tone, or sine-wave components, or it can be a completely random phenomenon made up of an infinite number of components, each having purely random amplitude and phase characteristics. Automobile exhaust noise, for example, contains pure-tone components related to the engine rotational speed, whereas an air jet hiss is a random noise.

The physical specification of such noises is given by their radiated intensity, frequency, and spatial distributions, as discussed elsewhere (*see* SOUND; WAVE MOTION). Random noises are usually described in terms of statistical values rather than in

closely related, on the average, are the present and future values of $v(t)$. For a stationary ergodic noise $R(\tau)$ is defined as the time average of a product of amplitudes

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} v(t)v(t+\tau)\,dt$$

The power spectral density $N(f)$ gives the distribution of average power in the noise waveform $v(t)$ as a *function of* frequency f, that is, $N(f)\,df$ is the average power a voltage $v(t)$ applied across a 1-ohm resistance would dissipate in the incremental frequency range f to $f + df$. $N(f)$ can be calculated as the Fourier transform of $R(\tau)$, or

$$N(f) = 2\int_{-\infty}^{\infty} R(\tau)\epsilon^{-j2\pi f\tau}\,d\tau \quad 0 \leq f \leq \infty$$

where $j = \sqrt{-1}$.

If a stationary noise waveform $v(t)$ with power spectral density $N(f)$ is passed through a linear time invariant filter with a transfer function $H(f)$, where $H(f)$ is the complex gain of the filter at the frequency f, the power spectral density of the output noise is given by $N(f)|H(f)|^2$. The output noise is also stationary, and if the input noise is gaussian, so is the output. However if gaussian noise is passed through a nonlinear device, such as a square-law detector or a limiter its statistics do not in general remain gaussian.

In analyzing the response of an electrical system to signals-plus-noise a quantity of interest is the ratio of signal-power to noise power, or, briefly, signal to noise ratio, at various points of the system. If $S(f)\,df$ is the signal power in the incremental frequency band df at f, the incremental signal-power-to-noise ratio is $S(f)/N(f)$. This ratio is unchanged by passing the signal plus noise through a linear filter. If $S(f)$ and $N(f)$ are each integrated over the total band of frequencies being passed, the ratio of the integrals is the total signal-power/noise-power ratio.

Thermal noise. In the narrow sense, thermal noise refers to the random voltage at the terminals of . . .

. . . voltage is

$$N(f) = 4kTR\gamma(f)$$

where k is Boltzmann's constant (1.38×10^{-23} joule/°C), R is the resistance in ohms of the resistor, T is its temperature in degrees Kelvin, and $\gamma(f)$ is given by

$$\gamma(f) = \frac{hf}{kT}\epsilon^{-(hf/kT)+1}$$

where h is Planck's constant (6.62×10^{-34} joule-sec).

At room temperature and even at microwave frequencies $\gamma(f)$ is very close to unity, so the noise voltage spectral density across a resistor is usually approximated by

$$N(f) = 4kTR$$

which is known as Nyquist's formula.

Nyquist's formula shows that the thermal noise voltage spectral density is constant (to very high frequencies). Such noise with a uniform distribution of power with frequency is called white noise.

Nyquist's formula remains valid if the simple resistor is replaced by a two-terminal *electrical* network with all resistance elements at the same temperature and with R standing for the input resistance at the terminals. Generalized forms are valid when the temperature is not constant throughout the network.

In a broader sense thermal noise refers to other kinds of noise of thermal origin. Important among these is antenna noise. Nyquist's formula can be used for a two-terminal system in which the terminals are connected to an antenna in a *perfectly* black chamber, the whole system in thermal equilibrium. The R of Nyquist's formula is then the *radiation* resistance of the antenna. In practice, Nyquist's formula is often used to define the *equivalent radiation* temperature of the portion of sky at which the antenna is looking, in terms of the radiation resistance and the measured noise.

Shot noise. In a vacuum tube in which electrons are emitted from a heated cathode, the electron *stream from cathode to anode* does not have uniform density, since the individual electrons are emitted randomly and with *random initial velocities*. Although the fluctuations in the electron stream are essentially thermal in *origin*, the *current fluctuations*, called the shot effect, are usually not classified as thermal noise. One important *difference between* the shot effect in tubes and thermal noise in resistors is that the former is not *present until a voltage difference* is applied between cathode and anode, whereas the latter is present in a quiescent circuit with no externally applied voltages (*see* SCHOTTKY EFFECT). Similarly in transistors, the fluctuations in the rate of flow of carrier electrons or holes is a type of noise not present in the quiescent device, and this is also often called shot effect.

A satisfactory mathematical model for the emission of electrons from a hot cathode yields the result that the number of electrons emitted in any interval of time follows the Poisson law:

$$\text{Probability } (K \text{ electrons are emitted in } \tau \text{ sec}) = \frac{(\bar{n}\tau)^K\epsilon^{-\bar{n}\tau}}{K!}$$

where $\bar{n}$ . . . per s . . . field . . . the e . . . emitted at the cathode travel to the anode and the fluctuations in the current induced in the anode circuit are proportional to the fluctuations in . . .

. . . the power spectral density of the shot noise is given approximately at

times, but during sunspot activity the intensity of its noise radiation is greatly increased.

Nonrandom noise. This type of noise is usually the result of radiation from other electric equipment, from unwanted coupling with other systems, or from spurious oscillations within an electrical circuit. *See* CROSSTALK; INTERFERENCE, ELECTRICAL.

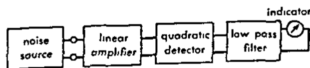
Noise measurement. The term noise measurement is applied to a wide range of measurements of random and nonrandom noise. It usually refers to the measurement of noise power averaged over some brief interval of time (called quadratic content). This kind of measurement may be made to check the noise level against which a system must operate, or it may be made to yield information about the physical world. In radio astronomy, for example, very delicate measurements are made of the noise radiation from particular planets, stars, or galaxies in order to gain information about their constitution and their relative velocity. In this application, the noise-measuring device is called a radiometer

Noise-power measurements are most conveniently made by amplifying the noise from the source in a linear amplifier and then using a quadratic detector followed by a low-pass filter and an indicating device to determine the average noise power.

The illustrated circuit measures noise power in the band of frequencies passed by the linear amplifier. The low-pass filter must have a narrower bandwidth than that of the linear amplifier to give an averaging of the fluctuating voltage from the detector (the low-pass filtering may occur naturally in the indicating device). In such a circuit, if the indicator responds linearly to its input voltage, its reading is proportional to the noise power Thermocouples or thermistors are good quadratic detectors. Electron-tube and crystal-diode detectors can be made to have nearly a quadratic response over a certain range. If the amplifier is nonlinear or the detector does not follow a quadratic law, the device will still indicate noise power, but its entire response curve must be calibrated.

Noise-power measuring devices may be either direct-reading or they may make comparisons between the noise from the source being measured and that from a calibrated source. *See* NOISE GENERATOR, ELECTRICAL.

Mathematical analysis of noise. The effects of nonrandom noise in an electric circuit or system can be determined mathematically in the same way that responses to signals in the circuit are determined. Random noise, being unpredictable and having only certain statistical properties, must be treated differently. Usually probabilistic methods are used.



Noise-power measurement

The application of probability theory to the analysis of random noise rests on the fact that for most mechanisms of noise generation (in particular, all those listed above except perhaps atmospherics) the resulting noise waveforms have statistical properties which remain invariant with time. These statistical properties may be deduced from an analysis of the generating mechanism or determined empirically by experiment.

One characterization of a random noise waveform $v(t)$ is in terms of the probabilities that the amplitude of the noise waveform will fall between any pair of specified values at specified instants of time. For example, the probabilities at a single instant t can be expressed in terms of a probability density function $f_t(x)$, defined by

$$f_t(x)\,dx = \text{probability that } (x < v(t) \leqq x + dx)$$

Then the probability that $v(t)$ lies between two values a and b is

$$\int_a^b f_t(x)\,dx$$

Most kinds of random noise, including particularly those which result from the summation of a great many microscopic effects, such as thermal noise and shot noise, have what are called gaussian or normal statistics. The term gaussian noise is used for such noise The probability density function for gaussian noise is

$$f_t(x) = \frac{1}{\sigma_t\sqrt{2\pi}}\,\epsilon^{-(x-m_t)^2/2\sigma_t^2}$$

where m_t and σ_t^2 are the parameters which determine $f_t(x)$ and are called the mean value (at time t) and the variance (at time t), respectively.

A random noise is said to be stationary if all its probability relations are unchanged by a translation in time. Thus, in particular, for a stationary noise $f_{t_1}(x) = f_{t_2}(x)$, for any times t_1 and t_2 Hence $f_t(x)$, m_t, and σ_t^2 do not depend on time and the subscript t may be dropped Thermal noise from a resistor at constant temperature, or shot noise from a diode under fixed operating conditions are examples of stationary noise. Usually a noise voltage $v(t)$ is stationary, and in addition has the property that the probability that $v(t)$ lies between any two values a and b at any time t is nearly equal to the fraction of time the noise voltage lies between these two values if a sufficiently long observation interval is recorded. This property, more precisely defined, is called ergodicity. For a stationary ergodic random noise the mean value m used above is just the average value of the noise over a long time interval and the variance σ^2 is the average of the square of the fluctuations about the mean

Other quantities that are useful in characterizing random noise are the covariance function $R(\tau)$ and (for a stationary noise) the power spectral density $N(f)$. The covariance function $R(\tau)$ of a noise waveform $v(t)$ gives a measure of ho

closely related, on the average, are the present and future values of $v(t)$. For a stationary ergodic noise $R(\tau)$ is defined as the time average of a product of amplitudes.

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} v(t)v(t+\tau)\,dt$$

The power spectral density $N(f)$ gives the distribution of average power in the noise waveform $v(t)$ as a *function of frequency* f; that is, $N(f)\;df$ is the average power a voltage $v(t)$ applied across a 1 ohm resistance would dissipate in the incremental frequency range f to $f + df$. $N(f)$ can be calculated as the Fourier transform of $R(\tau)$, or

$$N(f) = 2\int_{-\infty}^{\infty} R(\tau)\epsilon^{-j2\pi f\tau}\,d\tau \quad 0 \leq f \leq \infty$$

where $j = \sqrt{-1}$

If a stationary noise waveform $v(t)$ with power spectral density $N(f)$ is passed through a linear time invariant filter with a transfer function $H(f)$, where $H(f)$ is the complex gain of the filter at the frequency f, the power spectral density of the output noise is given by $N(f)|H(f)|^2$. The output noise is also stationary, and if the input noise is gaussian, so is the output. However if gaussian noise is passed through a nonlinear device, such as a square law detector or a limiter, its statistics do not in general remain gaussian.

In analyzing the response of an electrical system to signals-plus-noise a quantity of interest is the ratio of signal power to noise-power, or, briefly, signal-to-noise ratio, at various points of the system. If $S(f)$ ... [illegible]

unchanged by passing the signal plus noise through a linear filter. If $S(f)$ and $N(f)$ are each integrated over the total band of frequencies being passed the ratio of the integrals is the total signal-power/noise power ratio.

Thermal noise. In the narrow sense, thermal noise refers to the random voltage at the terminals of a ... [illegible] ... voltage is

$$N(f) = 4kTR\gamma(f)$$

where k is Boltzmann's constant (1.38×10^{-23} joule/°C), R is the resistance in ohms of the resistor, T is its temperature in degrees Kelvin, and $\gamma(f)$ is given by

$$\gamma(f) = \frac{hf}{kT}\epsilon^{-(hf/kT)+1}$$

where h is Planck's constant (6.62×10^{-34} joule sec).

At room temperature and even at microwave frequencies $\gamma(f)$ is very close to unity, so the noise voltage spectral density across a resistor is usually approximated by

$$N(f) = 4kTR$$

which is known as Nyquist's formula.

Nyquist's formula shows that the thermal noise voltage spectral density is constant (to very high frequencies). Such noise with a uniform distribution of power with frequency is called *white noise*.

Nyquist's formula remains valid if the simple resistor is replaced by a two-terminal electrical network with all resistance elements at the same temperature and with R standing for the input resistance at the terminals. Generalized forms are valid when the temperature is not constant throughout the network.

In a broader sense thermal noise refers to other kinds of noise of thermal origin. Important among these is antenna noise. Nyquist's formula can be used for a two-terminal system in which the terminals are connected to an antenna in a perfectly black chamber, the whole system in thermal equilibrium. The R of Nyquist's formula is then the *radiation* resistance of the antenna. In practice, Nyquist's formula is often used to define the equivalent radiation temperature of the portion of sky at which the antenna is looking, in terms of the radiation resistance and the measured noise.

Shot noise. In a vacuum tube in which electrons are emitted from a heated cathode, the electron *stream from cathode to anode* does not have uniform density, since the individual electrons are emitted randomly and with random initial velocities. *Although* the fluctuations in the *electron* stream are essentially thermal *in origin, the current fluctuations*, called the shot effect, are usually not classified as thermal noise. One important *difference* between the shot effect in tubes and thermal noise in resistors is that the former is not present until a voltage difference is applied between cathode and anode, whereas the latter is present in a quiescent circuit with no externally applied voltages (*see* SCHOTTKY EFFECT). *Similarly* in transistors, the fluctuations in the rate of flow of carrier electrons or *holes is a type of noise not present in the quiescent device*, and this is also often called shot effect.

A satisfactory mathematical model for the emission of electrons from a hot cathode yields the result that the number of electrons emitted in any interval of time follows the Poisson law.

$$\text{Probability } (K \text{ electrons are emitted in } \tau \text{ sec}) = \frac{(\bar{n}\tau)^K\epsilon^{-\bar{n}\tau}}{K!}$$

where $\bar{n}$ is the average number of electrons emitted per second. If a sufficiently great positive electric field is established in the cathode-anode space, all the electrons emitted at the cathode travel to the anode and the fluctuations in the current induced in the anode circuit are proportional to the ... [illegible]

... spectral ... shot noise is given approximately at

times, but during sunspot activity the intensity of its noise radiation is greatly increased.

Nonrandom noise. This type of noise is usually the result of radiation from other electric equipment, from unwanted coupling with other systems, or from spurious oscillations within an electrical circuit *See* CROSSTALK; INTERFERENCE, ELECTRICAL.

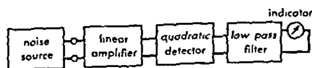
Noise measurement. The term noise measurement is applied to a wide range of measurements of random and nonrandom noise. It usually refers to the measurement of noise power averaged over some brief interval of time (called quadratic content). This kind of measurement may be made to check the noise level against which a system must operate, or it may be made to yield information about the physical world. In radio astronomy, for example, very delicate measurements are made of the noise radiation from particular planets, stars, or galaxies in order to gain information about their constitution and their relative velocity. In this application, the noise-measuring device is called a radiometer.

Noise-power measurements are most conveniently made by amplifying the noise from the source in a linear amplifier and then using a quadratic detector followed by a low-pass filter and an indicating device to determine the average noise power.

The illustrated circuit measures noise power in the band of frequencies passed by the linear amplifier. The low-pass filter must have a narrower bandwidth than that of the linear amplifier to give an averaging of the fluctuating voltage from the detector (the low-pass filtering may occur naturally in the indicating device). In such a circuit, if the indicator responds linearly to its input voltage, its reading is proportional to the noise power. Thermocouples or thermistors are good quadratic detectors Electron-tube and crystal-diode detectors can be made to have nearly a quadratic response over a certain range. If the amplifier is nonlinear or the detector does not follow a quadratic law, the device will still indicate noise power, but its entire response curve must be calibrated.

Noise-power measuring devices may be either direct-reading or they may make comparisons between the noise from the source being measured and that from a calibrated source. *See* NOISE GENERATOR, ELECTRICAL.

Mathematical analysis of noise. The effects of nonrandom noise in an electric circuit or system can be determined mathematically in the same way that responses to signals in the circuit are determined. Random noise, being unpredictable and having only certain statistical properties, must be



Noise-power measurement

treated differently. Usually probabilistic methods are used.

The application of probability theory to the analysis of random noise rests on the fact that for most mechanisms of noise generation (in particular, all those listed above except perhaps atmospherics) the resulting noise waveforms have statistical properties which remain invariant with time. These statistical properties may be deduced from an analysis of the generating mechanism or determined empirically by experiment.

One characterization of a random noise waveform $v(t)$ is in terms of the probabilities that the amplitude of the noise waveform will fall between any pair of specified values at specified instants of time. For example, the probabilities at a single instant t can be expressed in terms of a probability density function $f_t(x)$, defined by

$$f_t(x)\,dx = \text{probability that } (x < v(t) \leqq x + dx)$$

Then the probability that $v(t)$ lies between two values a and b is

$$\int_a^b f_t(x)\,dx$$

Most kinds of random noise, including particularly those which result from the summation of a great many microscopic effects, such as thermal noise and shot noise, have what are called gaussian or normal statistics. The term gaussian noise is used for such noise. The probability density function for gaussian noise is

$$f_t(x) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{-(x-m_t)^2/2\sigma_t^2}$$

where m_t and σ_t^2 are the parameters which determine $f_t(x)$ and are called the mean value (at time t) and the variance (at time t), respectively.

A random noise is said to be stationary if all its probability relations are unchanged by a translation in time. Thus, in particular, for a stationary noise $f_{t_1}(x) = f_{t_2}(x)$, for any times t_1 and t_2. Hence $f_t(x)$, m_t, and σ_t^2 do not depend on time and the subscript t may be dropped Thermal noise from a resistor at constant temperature, or shot noise from a diode under fixed operating conditions are examples of stationary noise. Usually a noise voltage $v(t)$ is stationary, and in addition has the property that the probability that $v(t)$ lies between any two values a and b at any time t is nearly equal to the fraction of time the noise voltage lies between these two values if a sufficiently long observation interval is recorded This property, more precisely defined, is called ergodicity For a stationary ergodic random noise the mean value m used above is just the average value of the noise over a long time interval and the variance σ^2 is the average of the square of the fluctuations about the mean.

Other quantities that are useful in characterizing *random noise are the covariance function $R(\tau)$* and (for a stationary noise) the power spectral density $N(f)$. The covariance function $R(\tau)$ of a noise waveform $v(t)$ gives a measure of how

(5) providing shielding walls or barriers, and (6) introducing mufflers into air-stream paths transmitting sound

Attenuation is achieved in structure-borne transmission paths by (1) mounting source or receiver on vibration isolation mounts, (2) increasing the vibration damping inherent to the transmission path by adding damping materials, and (3) providing discontinuities in the structural transmission path For a discussion of these last three techniques see VIBRATION DAMPING, VIBRATION ISOLATION.

Source placement Increasing the distance between source and receiver under free-field conditions will reduce the noise levels at the receiver in inverse proportion to the square of the distance For example. moving a source from 50 to 100 ft away from a receiver will reduce the received noise levels by 10 log $(100/50)^2$, or 6 decibels (db) See FREE FIELD, INVERSE-SQUARE LAW

Walls and enclosures. Surrounding a source by an impervious enclosure or separating a source and receiver located in the same room by interposing a wall between them to form two rooms will provide an amount of noise reduction which depends upon the transmission loss of the wall or enclosure, its surface area, and the amount of acoustical absorption in the receiving room. This noise reduction NR in decibels can be computed from the expression

$$NR = TL - 10 \log_{10}\left(\frac{1}{4} + \frac{S_w}{R}\right) \quad \text{db}$$

where NR = difference in sound pressure levels in decibels on the two sides of the wall (determined by measuring the sound pressure level on the primary side with a microphone that is moved around in the reverberant sound field and then subtracting from it the sound pressure level obtained with a microphone that is moved around in a region fairly near the surface on the secondary side)

TL = 10 times the logarithm to the base 10 of the ratio of the sound energy incident on the wall to the sound energy transmitted through the wall

S_w = area of the transmitting wall

R = room constant for receiving room = $[S\bar{\alpha}/(1-\bar{\alpha})]$, where S is the total area of the surfaces of the room on the secondary side and $\bar{\alpha}$ is the average absorption coefficient for the receiving room S must have the same dimensions as S_w

If the source and receiver are in the [illegible] or the [illegible] of [illegible] NR the [illegible] and other aspects of noise control within buildings, see ARCHITECTURAL ACOUSTICS; NOISE CONTROL IN BUILDINGS

Acoustical absorption. It can be seen from the preceding discussion that increasing the acoustical absorption in the receiving room increases the noise reduction. However, the amount of reduction obtained, in decibels, is only moderate. Because the TL will normally be 20–60 db or more, the effect of the wall is considerably more important than the effect of acoustical absorption in the receiving room

The noise reduction provided by acoustical absorption in an enclosure containing both source and receiver is dependent not only upon this absorption but on the directivity of the source and the distance of the receiver from the source as well The amount of noise reduction obtained by adding absorption is calculated from the expression

$$NR = 10 \log_{10}\left(\frac{Q}{4\pi r^2} + \frac{4}{R_1}\right) - 10 \log_{10}\left(\frac{Q}{4\pi r^2} + \frac{4}{R_2}\right)$$

where Q is the directivity index of the source, equal to unity for an omnidirectional (nondirectional) source; r is the distance from source to receiver, R_1 is the room constant, as previously defined, before adding acoustical absorption; and R_2 is the room constant after acoustical absorption is added. It should be observed that increasing the acoustical absorption in the enclosure by a factor of 10 will provide a noise reduction of 10 db, and that increasing the absorption by a factor of 100 provides a noise reduction of only 20 db.

Directional effects Relocating or rotating a source with strong directional characteristics will often provide substantial noise reduction For example [illegible] it is directed sideways to him. In a similar manner, substantial noise reduction of engine, compressor, furnace, and other exhausts can be achieved by directing them vertically upward rather than horizontally.

The amount of noise reduction obtained by this vertical directionality can be computed from the chart in Fig. 1, where the directivity index at 90° to the axis of the exhausting device is given for various frequency bands See DIRECTIVITY; SOUND.

Barriers Impervious barriers or partial partitions between noise sources and receivers will provide a moderate amount of noise reduction. The most effective use of a shielding barrier is obtained if it is located close to the source or receiver. The approximate noise reduction, in decibels, obtained by such a shield can be computed from the expression

$$NR = 10 + 10 \log \frac{2}{\lambda}\left\{R\left[\left(1 + \frac{H^2}{R^2}\right)^{1/2} - 1\right] + D\left[\left(1 + \frac{H^2}{D^2}\right)^{1/2} - 1\right]\right\}$$

where H is the height of the barrier above the line of sight between source and receiver, R is horizon-

low frequencies by the Schottky formula,

$$S(f) = 2eI$$

where e is the charge on an electron and I is the average anode current. This formula is valid if f is small compared to the reciprocal of the transit time of an electron from cathode to anode.

In most electron tubes, for example, receiving tubes operating with negative grid bias, a negative space charge develops near the cathode. This space charge has a damping effect on the shot effect fluctuations. The low-frequency spectral density is then approximately

$$S(f) = 2eI\Gamma^2$$

where Γ^2 depends on tube geometry, potentials, and cathode temperatures and has a value between 0 and 1.

Noise figure. The noise figure of an amplifier is a figure of merit which measures the noisiness of the amplifier relative to the noisiness of the source driving the amplifier. It is important in radio and radar receivers, for example, that the noise figure be low if weak signals in noise are to be received. The operating noise figure F_o, which is a function of frequency, may be defined to be the ratio of the available incremental signal-to-noise power ratio of the source (over a frequency band df) to the *available incremental signal-to-noise power ratio* of the amplifier output (over the same frequency band df) when the amplifier is driven by the source. If the source noise is ascribed to thermal origins, F_o depends on the effective noise temperature of the source A standard noise figure F is sometimes defined to be the operating noise figure with the source at 290°K. [W L R]

Bibliography: W B. Davenport, Jr, and W L. Root, *Introduction to the Theory of Random Signals and Noise*, 1958, A. van der Ziel, *Noise*, 1954

Noise control

The process of obtaining an acceptable noise environment for a particular observation point or receiver. An acceptable environment implies a balance between resultant noise levels and the attendant operational and economic limitations on the process whereby they are obtained *See* NOISE, ACOUSTIC.

Adequate noise control involves the entire system of noise source, transmission path, and receiver Measures can be taken to control any one or all of the three basic elements comprising the system The choice of noise-control techniques to be employed in any instance is largely dependent upon the relative economics of working on any particular element of the problem. For example, it is often much less expensive in initial cost and operational simplicity to supply factory workers with ear plugs than to redesign major pieces of machinery.

The first step in noise control is an analysis of the nature and extent of the problem. This analysis includes (1) the physical description of the intensity, frequency, and spatial distribution of the sound radiated by the major sources of noise contributing to the problem, and the paths by which the sound is transmitted to the receiver; (2) the designation of acceptable noise criteria for the receiver; and (3) the determination of the noise reduction required to achieve these criteria.

The description of the noise sources, at least on an engineering basis, can be based upon procedures developed from theoretical analyses and experimental measurements. These procedures are described for a number of sources later on in this article. The choice of adequate criteria, however, often depends largely upon the designer and the long-range cost. Significant progress has been made in analysis of human reactions to noise in an effort to obtain specifications involving, primarily, considerations of annoyance, interference with speech, and damage to hearing. A general discussion of these human criteria for noise control is presented later in this article. If the receiver is a particular piece of equipment which must be protected from excess noise, the criterion is usually chosen to avoid dynamic stress which would produce structural failure or malfunction In these problems, criteria must be derived individually for each case.

NOISE REDUCTION TECHNIQUES

Once an adequate specification has been made of the source and an acceptable criterion has been chosen, the amount of noise reduction required is simply determined as the difference, as a function of frequency, between the noise levels which would be produced at the receiver by the source in question and the noise level criterion selected. The designer now must exercise his ingenuity to produce this noise reduction by operating on the parts of the source-path-receiver system available to him.

Make-up of source. Noise control of the source is most often left to the manufacturer. Manufacturers of consumer goods such as household appliances have concentrated considerable effort on reducing the noise of their products Automobile, aircraft, and railroad-car manufacturers expend large amounts of money to reduce the noise of their equipment Lowered noise levels are often the selling point in consumer acceptance of an article.

Attenuation methods. On the other hand, because it is usually difficult for a user to alter the physical make-up of an article in order to lower the noise it produces, he must somehow reduce the noise by external means In general, the basic way in which noise control can be effected between the source and the receiver is to introduce attenuation in the air-borne and structure-borne paths between them Attenuation is generally provided in air-borne paths by one or more of the following approaches: (1) increasing the distance between source and receiver, (2) isolating source and receiver by impervious walls or enclosures, (3) providing acoustical absorption in spaces which contain both source and receiver, (4) making use of directional characteristics of source and receiver.

(5) providing shielding walls or barriers, and (6) introducing mufflers into air-stream paths transmitting sound.

Attenuation is achieved in structure-borne transmission paths by (1) mounting source or receiver on vibration isolation mounts, (2) increasing the vibration damping inherent to the transmission path by adding damping materials, and (3) providing discontinuities in the structural transmission path. For a discussion of these last three techniques *see* VIBRATION DAMPING; VIBRATION ISOLATION.

Source placement. Increasing the distance between source and receiver under free field conditions will reduce the noise levels at the receiver in inverse proportion to the square of the distance. For example, moving a source from 50 to 100 ft away from a receiver will reduce the received noise levels by 10 log $(100/50)^2$, or 6 decibels (db). *See* FREE FIELD; INVERSE-SQUARE LAW.

Walls and enclosures. Surrounding a source by an impervious enclosure or separating a source and receiver located in the same room by interposing a wall between them to form two rooms will provide an amount of noise reduction which depends upon the transmission loss of the wall or enclosure, its surface area, and the amount of acoustical absorption in the receiving room. This noise reduction *NR* in decibels can be computed from the expression

$$NR = TL - 10 \log_{10}\left(\frac{1}{4} + \frac{S_w}{R}\right) \quad \text{db}$$

where NR = difference in sound pressure levels in decibels on the two sides of the wall (determined by measuring the sound pressure level on the primary side with a microphone that is moved around in the reverberant sound field and then subtracting from it the sound pressure level obtained with a microphone that is moved around in a region fairly near the surface on the secondary side)

TL = 10 times the logarithm to the base 10 of the ratio of the sound energy incident on the wall to the sound energy transmitted through the wall

S_w = area of the transmitting wall

R = room constant for receiving room = $[S\bar{\alpha}/(1-\bar{\alpha})]$, where S is the total area of the surfaces of the room on the secondary side and $\bar{\alpha}$ is the average absorption coefficient for the receiving room. S must have the same dimensions as S_w

If the source and receiver are in the open and one or the other is being enclosed completely, the value of R approaches infinity, reducing the equation to $NR = TL + 6$ db. For the method of determining the TL of any given wall, and other aspects of noise control within buildings, *see* ARCHITECTURAL ACOUSTICS; NOISE CONTROL IN BUILDINGS.

Acoustical absorption. It can be seen from the preceding discussion that increasing the acoustical absorption in the receiving room increases the noise reduction. However, the amount of reduction obtained, in decibels, is only moderate. Because the TL will normally be 20–60 db or more, the effect of the wall is considerably more important than the effect of acoustical absorption in the receiving room.

The noise reduction provided by acoustical absorption in an enclosure containing both source and receiver is dependent not only upon this absorption but on the directivity of the source and the distance of the receiver from the source as well. The amount of noise reduction obtained by adding absorption is calculated from the expression

$$NR = 10 \log_{10}\left(\frac{Q}{4\pi r^2} + \frac{4}{R_1}\right) - 10 \log_{10}\left(\frac{Q}{4\pi r^2} + \frac{4}{R_2}\right)$$

where Q is the directivity index of the source, equal to unity for an omnidirectional (nondirectional) source; r is the distance from source to receiver; R_1 is the room constant, as previously defined, before adding acoustical absorption; and R_2 is the room constant after acoustical absorption is added. It should be observed that increasing the acoustical absorption in the enclosure by a factor of 10 will provide a noise reduction of 10 db, and that increasing the absorption by a factor of 100 provides a noise reduction of only 20 db.

Directional effects. Relocating or rotating a source with strong directional characteristics will often provide substantial noise reduction. For example, a small air jet used to [illegible] will produce [illegible] if its axis is [illegible] with the nois[illegible] [illegible] if it is directed sideways to him. In a similar manner, substantial noise reduction of engine, compressor, furnace, and other [illegible] [illegible]g them [illegible] by this [illegible] computed from the chart in Fig. 1, where the directivity index at 90° to the axis of the exhausting device is given for various frequency bands. *See* DIRECTIVITY; SOUND.

Barriers. Impervious barriers or partial partitions between noise sources and receivers will provide a moderate amount of noise reduction. The most effective use of a shielding barrier is obtained if it is located close to the source or receiver. The approximate noise reduction, in decibels, obtained by such a shield can be computed from the expression

$$NR = 10 + 10 \log \frac{2}{\lambda}\left\{R\left[\left(1 + \frac{H^2}{R^2}\right)^{1/2} - 1\right] + D\left[\left(1 + \frac{H^2}{D^2}\right)^{1/2} - 1\right]\right\}$$

where H is the height of the barrier above the line of sight between source and receiver, R is horizon-

low frequencies by the Schottky formula,

$$S(f) = 2e\bar{I}$$

where e is the charge on an electron and $\bar{I}$ is the average anode current. This formula is valid if f is small compared to the reciprocal of the transit time of an electron from cathode to anode.

In most electron tubes, for example, receiving tubes operating with negative grid bias, a negative space charge develops near the cathode. This space charge has a damping effect on the shot effect fluctuations. The low-frequency spectral density is then approximately

$$S(f) = 2e\bar{I}\Gamma^2$$

where Γ^2 depends on tube geometry, potentials, and cathode temperatures and has a value between 0 and 1.

Noise figure. The noise figure of an amplifier is a figure of merit which measures the noisiness of the amplifier relative to the noisiness of the source driving the amplifier. It is important in radio and radar receivers, for example, that the noise figure be low if weak signals in noise are to be received. The operating noise figure F_o, which is a function of frequency, may be defined to be the ratio of the available incremental signal-to-noise power ratio of the source (over a frequency band df) to the available incremental signal-to-noise power ratio of the amplifier output (over the same frequency band df) when the amplifier is driven by the source. If the source noise is ascribed to thermal origins, F_o depends on the effective noise temperature of the source. A standard noise figure F is sometimes defined to be the operating noise figure with the source at 290°K. [W.L.R.]

Bibliography: W. B. Davenport, Jr., and W. L. Root, *Introduction to the Theory of Random Signals and Noise*, 1958; A. van der Ziel, *Noise*, 1954

Noise control

The process of obtaining an acceptable noise environment for a particular observation point or receiver. An acceptable environment implies a balance between resultant noise levels and the attendant operational and economic limitations on the process whereby they are obtained. *See* NOISE, ACOUSTIC.

Adequate noise control involves the entire system of noise source, transmission path, and receiver. Measures can be taken to control any one or all of the three basic elements comprising the system. The choice of noise-control techniques to be employed in any instance is largely dependent upon the relative economics of working on any particular element of the problem. For example, it is often much less expensive in initial cost and operational simplicity to supply factory workers with ear plugs than to redesign major pieces of machinery.

The first step in noise control is an analysis of the nature and extent of the problem. This analysis includes (1) the physical description of the intensity, frequency, and spatial distribution of the sound radiated by the major sources of noise contributing to the problem, and the paths by which the sound is transmitted to the receiver; (2) the designation of acceptable noise criteria for the receiver; and (3) the determination of the noise reduction required to achieve these criteria.

The description of the noise sources, at least on an engineering basis, can be based upon procedures developed from theoretical analyses and experimental measurements. These procedures are described for a number of sources later on in this article. The choice of adequate criteria, however, often depends largely upon the designer and the long-range cost. Significant progress has been made in analysis of human reactions to noise in an effort to obtain specifications involving, primarily, considerations of annoyance, interference with speech, and damage to hearing. A general discussion of these human criteria for noise control is presented later in this article. If the receiver is a particular piece of equipment which must be protected from excess noise, the criterion is usually chosen to avoid dynamic stress which would produce structural failure or malfunction. In these problems, criteria must be derived individually for each case.

NOISE REDUCTION TECHNIQUES

Once an adequate specification has been made of the source and an acceptable criterion has been chosen, the amount of noise reduction required is simply determined as the difference, as a function of frequency, between the noise levels which would be produced at the receiver by the source in question and the noise level criterion selected. The designer now must exercise his ingenuity to produce this noise reduction by operating on the parts of the source-path-receiver system available to him.

Make-up of source. Noise control of the source is most often left to the manufacturer. Manufacturers of consumer goods such as household appliances have concentrated considerable effort on reducing the noise of their products. Automobile, aircraft, and railroad-car manufacturers expend large amounts of money to reduce the noise of their equipment. Lowered noise levels are often the selling point in consumer acceptance of an article.

Attenuation methods. On the other hand, because it is usually difficult for a user to alter the physical make-up of an article in order to lower the noise it produces, he must somehow reduce the noise by external means. In general, the basic way in which noise control can be effected between the source and the receiver is to introduce attenuation in the air-borne and structure-borne paths between them. Attenuation is generally provided in air-borne paths by one or more of the following approaches: (1) increasing the distance between source and receiver, (2) isolating source and receiver by impervious walls or enclosures, (3) providing acoustical absorption in spaces which contain both source and receiver, (4) making use of directional characteristics of source and receiver.

rotational speed Major pressure pulsations occur at the fundamental firing frequency f of the engine and at harmonics of f For a 4-cycle engine, $f = ns/120$ where n is the number of cylinders, s is the crankshaft speed in rpm, and f is in cps. At the normal operating speeds of automobile engines, effective reactive mufflers can be designed to reduce by substantial amounts the pulsating pressures causing intake and exhaust noise. Effective vibration mounting of the engine itself has resulted in low transmission of vibration induced by the engine into the body structure.

Gear noise from the transmission and differential of an automobile is largely a function of the quality of the gears and the amount of structural isolation which can be achieved in mounting the cases on the chassis Gear noise in general is described later.

A major contribution to noise control inside automobiles has been the use of mastic damping materials, such as body undercoating, to reduce vibration of sheet-metal components, and the use of acoustically absorbing materials to reduce the reverberation inside the passenger compartment

Noise produced by tires has been reduced substantially by redesign of tread patterns and methods of bonding the cord structures to the tread Tire thump, on the other hand, is still a major problem Reduction of this noise depends upon improved shock absorbers and spring suspensions of the running gear from the automobile body and upon even more effective damping of vibrations of the body structure than is presently available.

The high speeds of which modern automobiles are capable results in high noise levels from wind turbulence The turbulence noise associated with automobiles, like that associated with aircraft, increases with the fourth power of speed Reduction of this noise is obtained by aerodynamic shaping of the body and the minimizing of protuberances which can accentuate turbulence in the slipstream of air past the body

Compressor noise. In many industrial facilities, compressors are often the cause of major noise problems. Three types of compressor are found in common usage: axial, centrifugal, and reciprocating Each of these units produces noise by causing an alternating pressure to be developed at both its inlet and outlet openings. Theoretical analyses of the noise produced by such devices are useful primarily in estimating various physical models for predicting which operating parameters of the compressor are important in determining its noise output Empirical evaluations of compressor noise output are generally more useful in noise-control problems.

Axial and centrifugal compressors. The over-all sound power level PWL in decibels re (relative to) 10^{-13} watt produced by the discharge of axial and centrifugal compressors [illegible] by the following expression:

$$PWL = 20 \log \text{hp} + 40 \log v_t - 46$$

The sound power radiated from the intake of an axial compressor is essentially equal to that radiated from the discharge. The sound power radiated from the intake of a centrifugal compressor, however, [illegible] noise over a broad range of frequencies, with the predominant power occurring at frequencies associated with blade and impeller passage rates. The peak frequency output of axial compressors is usually at higher frequencies than those produced by centrifugal [illegible] Fig. 2.

Reciprocating compressors Reciprocating compressors produce a series of pressure pulses at their intakes and exhausts at the fundamental and harmonic frequencies of the rotational speed of

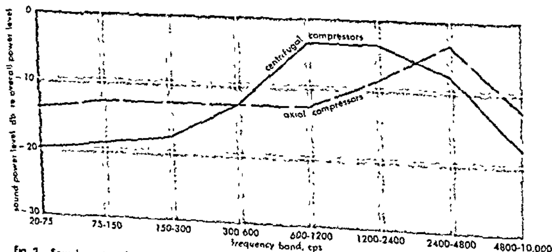


Fig 2. Sound spectra of axial and centrifugal compressors

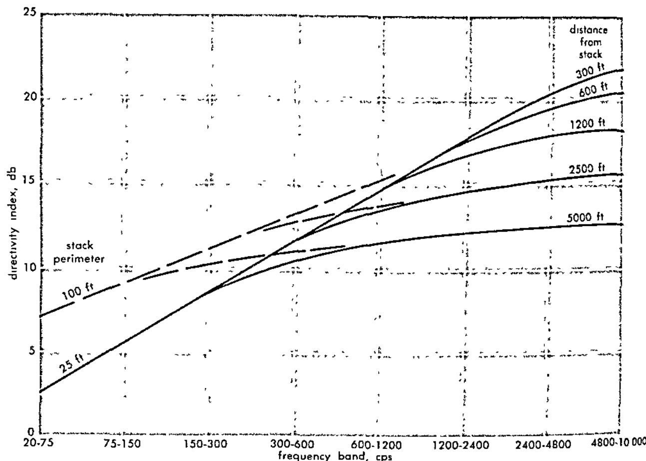


Fig 1 Directivity of a vertical stack

tal distance from source to barrier, D is horizontal distance from receiver to barrier, and λ is wavelength of sound. If D is much larger than R and R is larger than or equal to H, then the expression is approximately $NR = 10 + 10 \log (H^2/\lambda R)$.

Mufflers Many problems in noise control involve the reduction of sound being propagated in a pipe or duct along with the flow of a liquid or gas Engine exhausts, compressor intakes, ventilating systems, pumps, and other mechanical devices provide examples of this type of noise problem. The most effective noise control in these cases is often the addition of a structure to the system which will permit the fluid to flow but will attenuate sound. For a discussion of types of devices that perform this function *see* MUFFLER.

MAJOR NOISE SOURCES

The general discussion given in the preceding paragraphs outlines the basic approaches to identification and solution of a noise-control problem One of the primary requirements given was the specification of the physical properties of the noise source. The following paragraphs outline the noise characteristics of a number of general types of source and suggest means, where possible, for estimating these characteristics from the mechanical properties of the device.

Aircraft noise. Some of the major noise-control problems in present-day society are those produced by modern commercial and military aircraft. Noise in aircraft is a problem to both the passengers and to observers on the ground. Interior noise control on aircraft is obtained by introducing acoustical absorption in the cabin spaces and by providing adequate transmission loss through the cabin side walls and windows In order to provide lightweight structures, great ingenuity has been applied in the development of multiple-layer structures of thin solid materials interspersed between blankets of absorbing materials. These composite structures provide considerably higher transmission loss than an equivalent weight of solid material alone

Measures to control exterior noise produced by aircraft are primarily operational in nature Flight paths and power settings can be prescribed to minimize noise, and ground operations can be oriented to take maximum advantage of directional characteristics of noise radiation patterns. Mufflers and engine test cells are often employed to reduce noise from maintenance operations For a discussion of the mechanisms of noise production in aircraft, *see* AIRCRAFT NOISE

Automobile noise. The noise from automotive vehicles is produced primarily from engine intake and exhaust, vibration transmitted to the chassis from the motor, transmission and differential gears, tire treads, tires thumping against road irregularities, and wind turbulence about body protrusions

Intake and exhaust noise are primarily pulsating pressure phenomena associated [illegible] the engine

he noise spectrum can be estimated as being contained in an envelope where all octave bands above the band including the lowest gear contact frequency have the value just computed, and all octave bands below this band slope at a rate of +3 db/octave.

Bearing noise. Vibration in the bearing introduced by unbalanced loads is the primary cause of bearing noise These loads may be introduced by eccentricities in the bearing or shaft, overloads, local variations in frictional forces, and defects or wear. Noise control for bearings usually involves modifying the design by varying rotational speeds, changing lubricants, using larger or different types of bearings, and altering clearances. Reduction of noise radiated by bearings can also be achieved by increased damping of structures in which the bearing is mounted, vibration isolation, and exterior enclosures. No reliable means exists for predicting in advance the actual noise power that most bearing installations will produce

Rail transportation noise. The noises associated with railroad trains, subways, and trolley cars result primarily from impact of the wheels with joints in the steel rails and the scraping of the wheels on the sides of the rails in going around curves Noise-control measures to reduce these noises inside trains include better vibration isolation of passenger compartments and more skillfully designed wall, floor, and roof structures to provide increased noise reduction of air-borne sounds

Significant improvement in reducing rail noises has been accomplished by welding smooth junctions between connecting rails instead of bolting them together and by placing rubber tires on the passenger-car wheels [W J C]

Bibliography· L. L Beranek (ed), *Noise Reduction*, 1960, C Harris (ed) *Handbook of Noise Control*, 1957

HUMAN CRITERIA FOR NOISE CONTROL

In acoustics noise is defined as unwanted sound It may be unwanted because it causes general feelings of annoyance interferes with or prevents the hearing of speech or music or other wanted acoustical signals, fatigues and damages the auditory mechanism, and affects the performance of people at work Indirectly these noise problems can have serious social and economic effects upon communities

Noise is increasingly prevalent in modern-day living both at work and in the home The most common sources of noise are power driven vehicles and machines In order for work spaces and living environments to be acceptable to people, noise must be controlled or reduced The control of noise is usually expensive and for the process to be reasonably efficient it is necessary to establish criteria or standards which specify maximum limits for the noise

Some of the effects of noise on man depend upon the circumstances of a given situation and the attitudes of each listener. For example, an unexpected sound may cause fright reactions which diminish as the sound is repeated. Some persons respond to a sound differently from other persons The diverse effects of noise and variations in response in different listeners make it impossible to specify a general criterion for tolerable noise levels. However, criteria have been established for certain specific effects of noise on man.

Masking of speech. When two sounds contain the same or nearly the same frequency components, a person will hear the more intense sound but will be unable to detect the presence of the other. This phenomenon is called *auditory masking* Accordingly, speech signals or other wanted acoustic signals may be made inaudible by a noise that falls in roughly the same frequency range and is more intense than the wanted signal. *See* MASKING (SOUND).

Regardless of how many times a noise is repeated or how experienced an individual becomes to listening in noisy environments, the masking effects of the noise will always be the same Typically, speech signals have an intensity at the position of a listener's ear of about 65 db re 0 0002 dyne/cm². Many noises in the home, in industry, and in offices contain frequency components that are more intense and in the same frequency region as those present in speech. Such noises mask speech. Normal speech varies greatly in intensity from instant to instant. For this reason, a masking noise may mask the speech signal either partially or completely

Speech-interference level. Many of the complaints people make about noise stem from its interference with direct person-to-person speech, telephone communications, and the sound of radios and television sets. The amount of interference with speech communications can be estimated by taking the arithmetic average of the sound pressure levels in the octave bands 600–1200 cps, 1200–2400 cps, and 2400–4800 cps This average level, in decibels, is called the speech-interference level. This value has been related to the effectiveness of speech communications under various environmental conditions Only broad-band noises (most aircraft, factory, and some office noises fall in this category) should be evaluated by the speech-interference level, which does not properly evaluate the interference effects of narrow bands of noise.

Articulation and intelligibility tests For many noise-control purposes, the speech-interference level is a sufficiently accurate measure of the effect of a noise on speech communication However, the masking of speech by noise can be more precisely determined by use of the articulation index or by speech intelligibility tests than by the speech-interference level. The articulation index is determined by finding the difference between the intensity of the speech signal and the noise at 20 selected points along the frequency scale Speech intelligibility testing is conducted with standardized nonsense syllables, words or sentence tests,

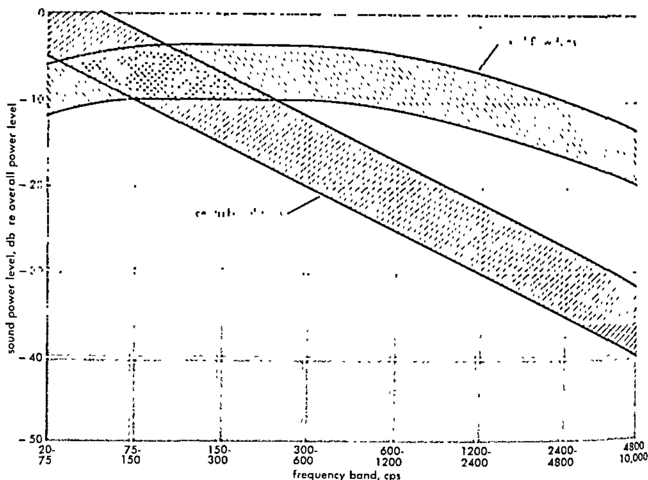


Fig 3 Sound spectra of axial and centrifugal fans.

the compressor in cps The sound power level PWL_f in decibels re 10^{-13} watt produced at the fundamental frequency of either the intake or exhaust of the compressor can be estimated from the respective intake or exhaust cylinder displacement D in cubic inches and rotational speed f in cps from the expression

$$PWL_f = 43 + 20 \log D + 20 \log f$$

The spectrum for such a compressor can be estimated by using the PWL_f as a base, and by assuming that harmonics will occur in the relationship $PWL_n = PWL_f - 6n$, where $n = 1, 2, 3, \ldots$. Octave frequency band values can be determined by summing the contributions from the individual harmonics lying in each frequency band in question.

Fan noise. Fans used in ventilating systems, air conditioners, automobiles, and forges produce noise in a fashion similar to that of aircraft propellers The movement of air is the principal job of the fan, however, rather than the production of thrust for propulsion of its mount. Fan manufacturers have expended large efforts in designing blade shapes and orientations which are efficient air movers but produce minimum noise. The variation in the sound power produced by any given fan from the average of fans in its size and performance range can easily amount to ± 5 db. The average sound power level of any given group of fans can be approximated, however, by the expression $PWL = 100 + 10 \log \text{hp}$, where PWL is in decibels re 10^{-13} watt and hp is the rated horsepower of the fan motor. The approximate spectrum shape of the noise from axial and centrifugal fans is shown in Fig. 3

Gear noise. The noise produced by gears is basically a function of the type of gear train, the machining tolerance employed, the speed of rotation, the number of teeth, the contact velocities, the horsepower transmitted, and the type of enclosure High machining tolerances, low frictional losses, and internal damping of both gears and housing all tend to reduce gear noise. Helical and hypoid gears tend to produce less noise than spur gears.

Noise control for gear trains usually involves the reduction of structurally transmitted noise by the use of resilient couplings and vibration mounts and the reduction of air-borne noise by additional damping of the gear enclosure or the use of an additional enclosure around the gear box.

The sound power of any given gear train is difficult to estimate in advance An estimate of the sound power level generated by spur gear trains of commercial grades can be made from the expression $PWL = 35 + 10 \log \text{hp} + 20 \log \text{rpm}$, where PWL is the sound power level in decibels re 10^{-13} watt in the octave frequency band containing the lowest gear contact frequency, hp is the horsepower transmitted by the gears, and rpm is the rotational speed per minute of the l[illegible]st-speed ge[illegible]

The noise spectrum can be estimated as being contained in an envelope where all octave bands above the band including the lowest gear contact frequency have the value just computed, and all octave bands below this band slope at a rate of +3 db/octave.

Bearing noise. Vibration in the bearing introduced by unbalanced loads is the primary cause of bearing noise. These loads may be introduced by eccentricities in the bearing or shaft, overloads, local variations in frictional forces, and defects or wear. Noise control for bearings usually involves modifying the design by varying rotational speeds, changing lubricants, using larger or different

[illegible]

and exterior enclosures No reliable means exists for predicting in advance the actual *noise power* that most bearing installations will produce

Rail-transportation noise. The noises associated with railroad trains, subways, and trolley cars result primarily from impact of the wheels with joints in the steel rails and the scraping of the wheels on the sides of the rails in going around curves Noise-control measures to *reduce these* noises inside trains include better vibration isolation of passenger compartments and more skillfully designed wall, floor, and roof structures to *provide* increased noise reduction of air-borne sounds

Significant improvement in reducing *rail* noises has been *accomplished* by welding smooth junctions between connecting *rails* instead of bolting them together and by placing rubber tires on the passenger-car wheels [W.J.G.]

Bibliography L. L Beranek (ed), *Noise Reduction*, 1960, C Harris (ed), *Handbook of Noise Control*, 1957.

HUMAN CRITERIA FOR NOISE CONTROL

In acoustics, noise is defined [illegible] It may [illegible] ings of [illegible] hearing [illegible] of music or other wanted acoustical signals, fatigues and damages the auditory mechanism, and affects the performance of people at work. Indirectly these noise problems can have serious *social* and economic effects upon communities

Noise is increasingly prevalent in modern-day living both at work and in the home The most common sources of *noise* are power-driven vehicles and machines In order for work spaces and living environments to be acceptable to people, noise must be controlled or *reduced* The control of noise is usually expensive and for the process to be reasonably efficient it is necessary to establish criteria or standards which specify maximum limits *for the* noise

Some of the effects of noise on man depend upon the circumstances of a given situation and the attitudes of each listener. For example, an unexpected sound may cause fright reactions which diminish as the sound is repeated. Some persons respond to a sound differently from other persons. The diverse effects of noise and variations in response in different listeners make it impossible to specify a general criterion for tolerable noise levels. However, criteria have been established for certain specific effects of noise on man

Masking of speech. When two sounds contain the same or nearly the same frequency components, a person will hear the more intense sound but will be unable *to detect the presence of the* other. This phenomenon is called *auditory masking* Accordingly, speech signals or other wanted acoustic signals may be made inaudible by a noise that falls in roughly the *same frequency* range and is more intense than the wanted signal. *See* MASKING (SOUND).

Regardless of how many times a noise is repeated or how *experienced* an individual becomes to listening in noisy *environments, the masking effects* of the noise will always be *the same.* Typically, speech signals have an intensity at the position of a *listener's* ear of about 65 db re 0 0002 dyne/cm². Many *noises in the* home, in industry, *and in* offices contain frequency components that are more intense and in the same frequency region as those present in speech. Such noises mask speech Normal speech varies greatly in intensity from instant to instant For this reason, a *masking noise* may mask the speech signal either partially or completely.

Speech interference level. Many of the complaints people make about noise stem from its interference with direct *person-to-person* speech, telephone communications, and the sound of radios and television sets. The amount of interference with speech *communications* can be estimated by *taking the arithmetic* average of the sound pressure levels in the octave bands 600–1200 cps, 1200–2100 cps, and 2400–4800 cps *This* average level, in decibels, is *called* the speech interference level This value has been *related* to the effectiveness of speech communications under *various* environmental conditions Only broad-band noises (most aircraft, *factory,* and *some* office noises fall in this category) *should* be evaluated by the *speech-interference level,* which does not properly evaluate *the interference effects of narrow* bands of *noise*

Articulation and intelligibility tests. For many noise-control purposes, the speech-interference level is a sufficiently accurate measure of the effect of a noise on [illegible] masking of [illegible] determined [illegible] speech intelligibility tests than by the speech-interference level. The articulation *index* is determined by finding the *difference* between *the* intensity of the speech signal and the noise at 20 selected points along *the* frequency scale Speech intelligibility testing is *conducted* with standardized nonsense syllables, words or sentence tests,

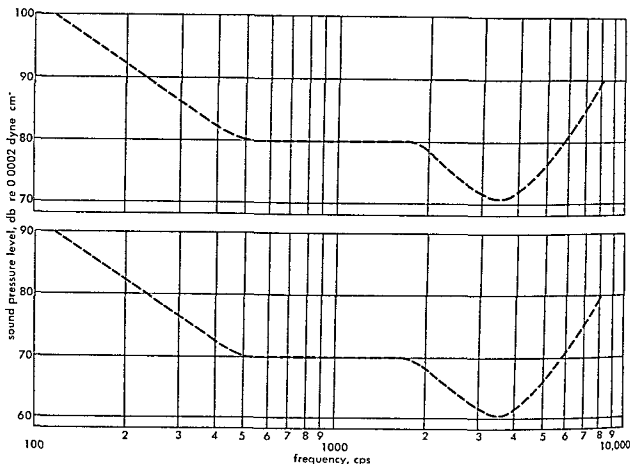


Fig 4 Proposed damage risk criteria for wide-band noise (upper curve) and for pure tones or critical bands of noise (lower curve).

and a trained crew of talkers and listeners. *See* PSYCHOACOUSTICS; SPEECH.

Hearing loss. An extremely important aspect of noise control is concerned with the prevention of damage to the hearing of people who are exposed to noise of high intensity. In these situations speech communication is usually unnecessary or is held to a minimum.

Hearing loss is usually measured by pure-tone audiometry, which measures the intensity in decibels of a pure tone that is just audible to a person in the quiet (*see* AUDIOMETRY). This process is repeated at a number of different frequencies, and the values thus obtained constitute an audiogram called the threshold of hearing. The normal threshold of hearing has been determined for persons of different age groups The difference between the normal threshold of hearing for a particular age group and the threshold of hearing for a given individual of a comparable age is the measure of the amount of hearing loss suffered by that individual.

It has been found that intense sound can cause either a temporary hearing loss from which the person recovers in time or a permanent loss from which there is only slight or no recovery.

The amount of hearing loss that is suffered because of exposure to intense noise is influenced by a number of factors. The following factors are important: (1) the frequency and intensity of the noise, (2) the bandwidth of the noise, (3) the duration of exposure during a single day, and (4) the number of years of working-day exposures

Damage risk criteria Two basic criteria have been deduced from research on hearing loss due to exposure to sound One criterion applies to sounds that are pure tones or to sounds that contain most of their energy in narrow, so-called critical bands. Critical bands vary somewhat in width as a function of frequency but are approximately 100 cps wide for most of the audible range of sound frequencies The second criterion applies to wide-band noise, in which the energy is spread over an octave or several octave bands. These two criteria are shown in Fig 4 That figure shows the maximum pressure level that a sound can possess if persons are not to suffer a possible hearing loss as the result of exposure to the sound The exposure may be for as long as 8 hr/day for a working life of 20–30 yrs.

The criteria in Fig. 4 are called damage risk for hearing. They are based primarily on industrial surveys and are designed to provide adequate protection for all people with normal ears. Many persons can be exposed to greater intensities than those shown in Fig. 4 without suffering any appreciable hearing loss.

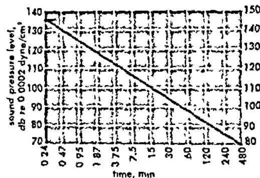


Fig. 5 Proposed maximum permissible sound pressure levels for pure tones or critical bands of noise (left-hand ordinate) and for octave bands of noise (right-hand ordinate) for frequencies of 500–1800 cps as a function of daily exposure time.

People can tolerate greater intensities than those shown in Fig. 4 when exposure to a sound or noise has a total duration of less than 8 hr/day. When the exposure time is reduced by a factor of 2, for example from 8 to 4 hr, the maximum permissible level is increased by 6 db, halving the exposure time again, from 4 to 2 hr, permits a further increase of 6 db in the maximum permissible sound pressure level. Figure 5 shows how the damage risk criteria for sounds above 300 cps change as a function of the duration of daily exposure time.

Equivalent exposure time. In industries and in the military services, persons are often exposed during a single day to noises of different intensities and durations. In these situations, the durations of exposure at each intensity are expressed in terms of equivalent exposure time. Equivalent exposure time is computed with reference to the level of a sound of the same frequency and bandwidth that is at the damage risk criterion for an 8-hr (480-min) exposure. For example, a 30-min exposure during a single day to an octave band of noise from 600 to 1200 cps at 86 db is equivalent to a 60-min exposure to a similar noise at a level of 80 db (the maximum level permitted for that octave band of noise for a 480-min period). The equivalent exposure time for this example is therefore 60 min. In order to determine the noise exposure experienced by a particular person, the actual durations and levels present during a typical working day are converted into equivalent exposure times. These equivalent exposure times are then added together; if the total exceeds 480 min it is concluded that the noise exposure is potentially dangerous to hearing. See DEAFNESS, EAR PROTECTORS

Noise and work performance. Many studies have been conducted on the effects of noise on work output in industry and upon the performance of psychological tests and motor-skill tasks in the laboratory. The results of these studies indicate that, in general, noise does not have an adverse effect upon work performance provided (1) speech communication is not involved in the work and (2) the workers have become adapted to the noise.

The following represent the conclusions of major studies conducted on this problem. The work performed in these studies did not require a significant degree of speech communication.

1. Initially workers are annoyed with and upset by the presence of intense industrial noise; however, after a few days or weeks they become adapted and do not object to the noise.

2. In some occupations, the reduction of noise results in an initial increase in work output. This initial increase tends to diminish with the passage of time.

3. In some tasks, for example, typing, work output is slightly greater in noise than in quiet. The noise apparently isolates the worker from auditory distractions and permits greater concentration.

4. Laboratory experiments in which subjects worked during alternating periods of noise and quiet show that subjects do as well on mental and motor tasks in the noise as in the quiet.

5. There is some indication that greater effort is expended when persons work in noise than in the quiet. However, with continued exposure and adaptation to the noise this difference in the amount of effort expended disappears.

These negative or equivocal findings indicate that noise-control criteria are not required because of any possible effects of noise on work performance. The application of speech-interference level, or equivalent, criteria where speech communication is an important part of the job will provide a noise environment that is acceptable for those work

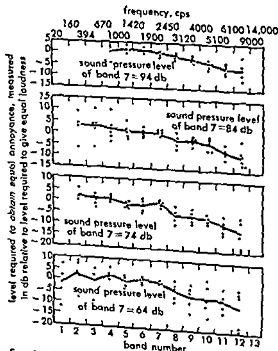


Fig. 6 Equal annoyance contours for bands of noise. (After K. D. Kryter)

spaces. The application of damage risk criteria where speech communication is not an important aspect of the job will provide a noise environment that is acceptable and which will not interfere with work output.

Annoyance feelings. Independently of other effects they may have, some sounds are judged by people to be inherently more annoying than other sounds. Other factors being equal, the higher the frequency components, or pitch, of a sound, the more annoying it tends to be. This fact was determined by asking people to adjust the intensity of sounds containing different frequency components until they thought that each sound would be as annoying as (or conversely, as acceptable as) each of the other sounds (see Fig. 6).

People judge with fair consistency how annoying, or how acceptable, they think one sound is relative to another sound. However, people do not consistently agree among themselves as to how tolerable, in an absolute sense, a given sound or noise is. Laboratory experiments and surveys of public opinion (for example, around airports and along heavily traveled highways) reveal that with continued exposure people become adapted in some respects to sounds of great intensities, intensities which initially may have been considered by them to be intolerable This adaptation progresses most rapidly when the listeners lose their fear of the sound, when they expect the sound as part of their normal environment, and when they learn to sleep in its presence

Largely because of these adaptation effects, it has not been possible to specify a criterion or standard for noise with respect to feelings of annoyance or interference with sleep. However, there are at least two major effects of noise to which man does not become adapted even though he is exposed repeatedly to the noise. These are interference with speech communication and damage to hearing.

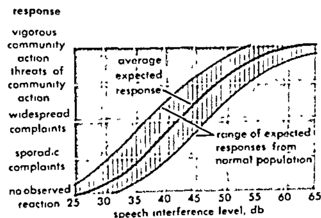


Fig. 7. Relation between response of residents in a neighborhood to the speech-interference level of the noise causing the response. (Modified from figure and method proposed by K. D Stevens, W A Rosenblith, and R H. Bolt)

Corrections in db to be applied to speech-interference level (SIL) noise rating*

Influencing factor	Possible conditions	Number of db to be added to SIL
Spectrum character	Continuous	0
	Pure-tone components	+5
Peak factor	Continuous	0
	Impulsive	+5
Repetitive character (20- to 30-sec exposures assumed)	1 exposure/min (or continuous)	0
	10–60 exposures/hr	−5
	1–10 exposures/hr	−10
	4–20 exposures/day	−15
	1–4 exposures/day	−20
	1 exposure/day	−25
Background noise	Very quiet suburban	+5
	Suburban	0
	Residential urban	−5
	Urban near some industry	−10
	Area of heavy industry	−15
Time of day	Daytime only	−5
	Nighttime	0
Adjustment to exposure	No previous exposure	0
	Considerable previous exposure	−5
	Extreme conditions of exposure	−10

* After L L Beranek, 1954

Community reaction. Noises from factories, passing aircraft, automobiles, and trucks are the cause of annoyance and complaints in a number of communities In some cases, legal action has been taken by private citizens and by municipalities to have the noise stopped as a public nuisance.

Surveys of this problem show a wide range in the responses made by people to noises found in communities. Figure 7 summarizes the general findings of these surveys In Fig 7, the response of people is plotted against the intensity of the noise in terms of its speech-interference level. The area covered by the curve in Fig. 7 is for typical conditions.

The results shown in Fig. 7 can be adjusted to make special allowance for the effects of certain environmental and noise conditions. Some of these factors, as shown in the accompanying table, tend to make a noise more or less acceptable than is indicated in Fig 7. The speech-interference level of a given environmental noise should be corrected in accordance with the table before being incorporated in Fig 7 to determine its probable effect upon a community.

It should be emphasized that for any individual community or neighborhood there may be factors present that cause the response to noise to be considerably different from that predicted by reference to the table and to Fig. 7. [K.D K.]

Bibliography: L. L. Beranek, *Acoustics*, 1954, L. L. Beranek, Revised criteria for noise in buildings, *Noise Control*, 3(1):19-27, 1957, C. M. Harris (ed), *Handbook of Noise Control*, 1957; K. D. Kryter, *The Effects of Noise on Man*, J Speech and Hearing Disorders, Monogr. suppl. 1, 195 ·

K. N. Stevens, W. A. Rosenblith, and R. H. Bolt, A community's reaction to noise: can it be forecast?, *Noise Control*, 1(1):63–71, 1955.

Noise control in buildings

The technology of obtaining an acceptable noise environment in various types of enclosures. Satisfactory noise control in buildings can be achieved most economically if it is given consideration in the planning stages of the building, for example, in the selection of the location of a building on its site, in the arrangement of rooms, corridors, and vestibules, and in the location of doors and windows.

Sources of noise in a building originate either in the air, or from direct impacts, or from a combination of both. The sounds which originate in the air are transmitted (1) along a continuous air path through openings such as ventilating ducts, open [illegible]

[illegible]

wall to the other.

Sound insulation. The methods of insulating solid-borne impact sounds are somewhat different from those of insulating air-borne sounds. A structure that is a very good insulator for one type may be a very poor insulator for the other. For example, a bare concrete slab 1 ft thick provides high insulation against air-borne sound, but it propagates impacts readily. However, some constructions provide excellent insulation for both air-borne and solid-borne sound.

Air-borne sound can be effectively controlled by the use of heavy walls and ceilings or by special types of multiple wall structures; solid-borne sound can be suppressed by discontinuities in the transmission path. The advantages of good insulation against both types of noise may be incorporated in discontinuous construction, that is, construction in which the rooms in a building are treated essentially as a suspended "box within a shell." The walls of a [illegible] floor, th [illegible] floor bu [illegible] or by a [illegible] ties between the walls and the continuous construction are avoided, but when necessary, resilient isolators are employed. The ceiling is suspended from the structural floor by resilient hangers to leave enough space above the false ceiling [illegible]

T [illegible] alm [illegible] tail [illegible] windows and doors should not form a rigid link between a detached room and the surrounding continuous construction [illegible] [illegible] supports [illegible] they penetrate walls, pipes should be isolated from the partitions by rubber, felt, or other compliant material. Care should be exercised to prevent cracks at these junctions and around doors, otherwise the insulation against airborne sounds will be reduced significantly.

Noise reduction by absorption. In addition to providing good air-borne and solid-borne sound insulation, it is necessary, in rooms where very low noise levels are required, to use a considerable amount of sound-absorptive material. The noise reduction provided by an increase in absorption in an enclosure can be estimated by the following equation. If the acoustic-power output of a noise source remains constant, and if the total absorption in the room is increased from a_1 to a_2, the reduction in noise level in decibels is given approximately by $10 \log (a_2/a_1)$. Thus if the absorption in a room is increased by a factor of 4, the average noise reduction will be about 6 db. (Note that reduction is different at different frequencies because the total absorption is a function of frequency.) This equation shows that the addition of absorptive materials in a room will provide substantial noise reduction in a room that is relatively bare. However, if the boundaries and furnishings of a room absorb much sound, the addition of absorptive treatment on the ceiling may not produce a significant reduction in noise level. In rooms that have high ceilings, treatment of the ceiling only may not yield satisfactory results, because sustained reflections may take place between the hard side walls. In such rooms some acoustical treatment should be applied to the side walls (for example, in the form of panels) as well as on the ceiling. Besides reducing the steady-state noise level, the addition of absorptive treatment also provides beneficial effects by reducing the reverberation time in the room and by localizing the source of noise to the area in which it originates. See Absorption (sound); Reverberation.

Acceptable noise levels. The accompanying table gives values of recommended acceptable average noise levels for unoccupied rooms with the ventilation system in operation. These values are

Acceptable average noise levels in unoccupied rooms*

Type of room	Recommended level, db re 0.0002 dyne/cm²
Radio, recording, and television studios	25–30
Music rooms	30–35
Legitimate theaters	30–35
Hospitals	35–40
Motion-picture theaters, auditoriums	35–40
Churches	35–40
Apartments, hotels, homes	35–45
Classrooms, lecture rooms	35–40
Conference rooms, small offices	40–45
Courtrooms	40–45
Private offices	40–45
Libraries	40–45
Large public offices, banks, stores	45–55
Restaurants	50–55

*After V. O. Knudsen and C. M. Harris. The levels given in this table are "weighted"; that is, they are the levels measured with a standard sound-level meter incorporating an "A" 40-db frequency-weighting network.

spaces The application of damage risk criteria where speech communication is not an important aspect of the job will provide a noise environment that is acceptable and which will not interfere with work output.

Annoyance feelings. Independently of other effects they may have, some sounds are judged by people to be inherently more annoying than other sounds. Other factors being equal, the higher the frequency components, or pitch, of a sound, the more annoying it tends to be. This fact was determined by asking people to adjust the intensity of sounds containing different frequency components until they thought that each sound would be as annoying as (or conversely, as acceptable as) each of the other sounds (see Fig 6).

People judge with fair consistency how annoying, or how acceptable, they think one sound is relative to another sound However, people do not consistently agree among themselves as to how tolerable, in an absolute sense, a given sound or noise is Laboratory experiments and surveys of public opinion (for example, around airports and along heavily traveled highways) reveal that with continued exposure people become adapted in some respects to sounds of great intensities, intensities which initially may have been considered by them to be intolerable. This adaptation progresses most rapidly when the listeners lose their fear of the sound, when they expect the sound as part of their normal environment, and when they learn to sleep in its presence.

Largely because of these adaptation effects, it has not been possible to specify a criterion or standard for noise with respect to feelings of annoyance or interference with sleep However, there are at least two major effects of noise to which man does not become adapted even though he is exposed repeatedly to the noise. These are interference with speech communication and damage to hearing.

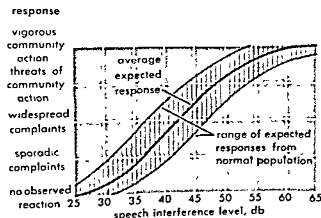


Fig. 7. Relation between response of residents in a neighborhood to the speech-interference level of the noise causing the response. *(Modified from figure and method proposed by K. D. Stevens, W. A Rosenblith, and R. H Bolt)*

Corrections in db to be applied to speech-interference level (SIL) noise rating*

Influencing factor	Possible conditions	Number of db to be added to SIL
Spectrum character	Continuous	0
	Pure-tone components	+5
Peak factor	Continuous	0
	Impulsive	+5
Repetitive character (20- to 30-sec exposures assumed)	1 exposure/min (or continuous)	0
	10–60 exposures/hr	−5
	1–10 exposures/hr	−10
	4–20 exposures/day	−15
	1–4 exposures/day	−20
	1 exposure/day	−25
Background noise	Very quiet suburban	+5
	Suburban	0
	Residential urban	−5
	Urban near some industry	−10
	Area of heavy industry	−15
Time of day	Daytime only	−5
	Nighttime	0
Adjustment to exposure	No previous exposure	0
	Considerable previous exposure	−5
	Extreme conditions of exposure	−10

* After L L Beranek, 1954

Community reaction. Noises from factories, passing aircraft, automobiles, and trucks are the cause of annoyance and complaints in a number of communities In some cases, legal action has been taken by private citizens and by municipalities to have the noise stopped as a public nuisance

Surveys of this problem show a wide range in the responses made by people to noises found in communities Figure 7 summarizes the general findings of these surveys. In Fig 7, the response of people is plotted against the intensity of the noise in terms of its speech-interference level. The area covered by the curve in Fig 7 is for typical conditions.

The results shown in Fig 7 can be adjusted to make special allowance for the effects of certain environmental and noise conditions Some of these factors, as shown in the accompanying table, tend to make a noise more or less acceptable than is indicated in Fig 7 The speech-interference level of a given environmental noise should be corrected in accordance with the table before being incorporated in Fig. 7 to determine its probable effect upon a community.

It should be emphasized that for any individual community or neighborhood there may be factors present that cause the response to noise to be considerably different from that predicted by reference to the table and to Fig. 7. [K.D K.]

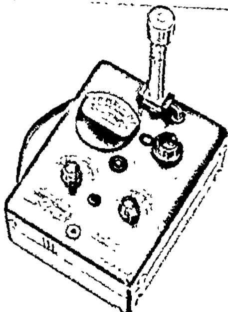
Bibliography: L. L Beranek, *Acoustics*, 1954; L. L. Beranek, Revised criteria for noise in buildings, *Noise Control*, 3(1):19–27, 1957; C. M. Harris (ed), *Handbook of Noise Control*, 1957; K. D Kryter, *The Effects of Noise on Man*, J. Speech and Hearing Disorders, Monogr. suppl. 1, 1950.

ments and techniques used for measuring noise are of fundamental importance in noise control. *See* NOISE, ACOUSTIC; NOISE CONTROL.

Instruments may be used to measure the amplitude of a sound as a function of time or frequency at any point in an acoustical noise field. Although limited noise measurements may be performed with mechanical devices, almost all noise measurements today are performed with the aid of electronic equipment. Noise measurement involves the use of an electroacoustic transducer (a microphone in air, a hydrophone in water) which transforms the sound pressure at the point of observation into a corresponding electrical signal. This elec- [illegible] *cating instrument*, such as a meter or an oscilloscope

For noise measurements in air, an instrument standardized by the American Standards Association is available for measuring the approximate rms amplitude of a noise on a weighted logarithmic scale This instrument is called a sound-level meter (SLM), and a reading in decibels (db) obtained on it is called a sound level. A standard instrument also exists for filtering a microphone signal into octave bands of frequency, and indicating the amplitude in each band on a logarithmic scale. This instrument is called an octave-band analyzer (OBA)

Specialized noise measurements, such as the determination of the peak or duration of a transient, or the computation of correlation functions, are occasionally performed These measurements require the use of peak reading meters, oscillographs or oscilloscopes, time-delay machines,



1 1 General Radio Type 1551-B sound-level meter. Company)

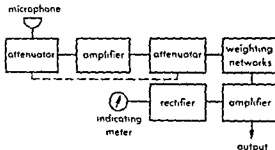


Fig 2 Simplified block diagram of sound-level meter.

matched filter sets, and the like. Often, noise measurements are facilitated by the use of a magnetic tape recorder which preserves the electrical signal produced by the microphone for subsequent detailed analysis.

Noise measurements are usually required to determine the properties of a noise field so that the optimum means for controlling this noise field may be determined. In general, it is possible to make noise measurements only at particular points in a noise field. Thus many measurements may be required to determine adequately the distribution of noise in a space.

This article gives detailed information on the sound level meter and on various types of sound analyzers For information on other devices and techniques commonly employed in the measurement of sound, *see* HYDROPHONE; MAGNETIC RECORDING, MICROPHONE; OSCILLOSCOPE, CATHODE-RAY; RECORDING INSTRUMENTS, GRAPHIC; SOUND; VOLTMETER.

Sound-level meter. This is an instrument which measures the approximate rms amplitude of a sound weighted according to frequency content and which meets the requirements of the American Standards Association. The sound-level meter was originally standardized to indicate sound level, a single number in db giving the total sound-pressure level weighted by an approximation to the loudness-level sensitivity of the human ear for pure tones [illegible] amplifier between the microphone and some other analyzing instrument.

A typical sound-level meter (Fig 1) consists of a microphone (usually of the piezoelectric type), an amplifier with a calibrated logarithmic attenuator, a set of frequency response (weighting) networks, and an indicating meter with a logarithmic scale (Fig. 2). The electrical signal produced by the microphone in the presence of a sound is read on the meter after having been weighted with respect to frequency content by the weighting networks The sound level is the sum of the meter reading and the attenuator setting.

The sound-level meter is basically a device for field use, and as such it is generally self-contained

used for design purposes, for example, in computing the amount of over-all noise insulation that should be provided for a room. They hold for typical room-noise spectra. Although even lower noise levels than those which are listed may provide some advantage under certain circumstances, and may be desirable if cost is not a factor, this table gives values which represent a combination of acceptability and economic practicality. For certain types of rooms the values which are recommended are lower than those which are commonly found.

Control of solid-borne noise. Mechanical energy, for example an impact, may be imparted directly to a building structure and then transmitted with little attenuation to another part of a building where a surface is forced into vibration, thereby radiating noise. Such solid-borne noise should be suppressed at its source wherever it is practical to do so Examples of control methods include the use of heavy carpeting, cork tile, or linoleum on felt to reduce impact transmission to the floor

In rating various types of floor constructions regarding their solid-borne sound insulation, a device called a tapping machine is used. This device can produce steady impacts (10 per sec) on the floor. The steady-state sound pressure levels in octave (or fractional-octave) bands produced by this source are measured in the room below or in some other room in the building. From these data the impact-sound insulation is determined. This quantity represents the improvement in insulation in decibels that the test floor construction provides over another one which is arbitrarily selected as a standard of reference. A floating floor construction usually provides high values of impact-sound insulation. Further improvement is obtained with a resilient floor finish which reduces the noise in the room where the impact is produced, thereby lowering the resulting noise level elsewhere in the building. *See* ARCHITECTURAL ACOUSTICS; NOISE CONTROL; VIBRATION ISOLATION. [C.M.H.]

Bibliography. V. O Knudsen and C. M. Harris, *Acoustical Designing in Architecture*, 1950; P. H. Parkin and H. R Humphreys, *Acoustics, Noise and Buildings*, 1958.

Noise filter, radio

A filter used in radio communications receivers to reduce noise. Usually it is an auxiliary low-pass filter which can be switched in or out of the audio system. The noise filter may also be equipped with a switch to vary the effective receiving bandwidth to meet the existing conditions of interfering noise. The tone control of a radio or record player can act as a noise filter, as when high-frequencies are cut down to reduce record noise. A band-pass filter may also be used, if the noise has a band-limited spectrum. *See* FILTER, ELECTRIC. [W.R L.]

Noise generator, electrical

A device which produces (usually random) electrical noise for use in electrical measurements. Electrical noise generators are commonly used in measuring the noise figure of a radio receiver or other amplifier. They are also used in other tests of the response of an electrical system to random noise, and in measurements of noise intensity. *See* NOISE, ELECTRICAL.

Some standard types of noise generator are: hot-wire, diode, gas-discharge tube, and klystron A hot-wire noise generator is commonly the filament of a lamp heated by a direct current. The filament is connected across the terminals where the noise is to be introduced, for example, the antenna terminals of a radio receiver. The noise generated by the filament is thermal, so its intensity N can be calculated from the Nyquist formula $N = 4kTR$, where T is temperature (°K), R is resistance of the filament, and k is Boltzmann's constant. A diode noise generator relies on the shot effect to produce noise. At frequencies appreciably less than the reciprocal of the transit time, the noise intensity N can be calculated from the Schottky formula $N = 2e\bar{I}$ if, as is customary in this application, the anode current is emission-limited. In this formula e is the charge of an electron and $\bar{I}$ is the average anode current. A gas-discharge noise generator is usually a fluorescent light tube enclosed in a wave guide. The mechanism of noise production is essentially thermal; the electrons in the gas discharge acquire high random velocities, corresponding to a high equivalent noise temperature. This equivalent noise temperature varies with the gas in the tube, but does not depend very much on tube dimensions or on the discharge current. A reflex klystron, with reflector grid connected to the cavity to prevent oscillation, generates noise because of shot effect in the cathode current.

It is convenient if the noise generated by a noise generator has a nearly constant intensity; certainly it must not fall off very much, over the entire frequency range of operation. Thermal noise sources do potentially provide such a flat spectrum of noise over all radio frequencies, being frequency-limited only by capacitances and inductances inherent in the source device and its circuitry. Diodes, however, are ultimately limited to frequencies of the order of the reciprocal of the transit time. In practice, special diodes can be used up to a few hundred megacycles. For the measurement of noise figure at microwave frequencies, where most amplifiers (triodes, traveling-wave tubes, and klystrons) have a relatively high noise figure (as high as 20 db above basic thermal noise), gas-discharge tubes are preferable to hot-wire noise sources because they produce more available noise power. Klystrons are also suitable noise generators at microwave frequencies, but they are not absolute standards and require calibration [W.L.R]

Bibliography: A. van der Ziel, *Noise*, 1954

Noise measurement

The process of quantitatively determining one or more of the properties of acoustic noise. In noise control research, knowledge of the physical properties of the undesirable sound constitutes the initial step toward an understanding hat shoul done educe or elimin he

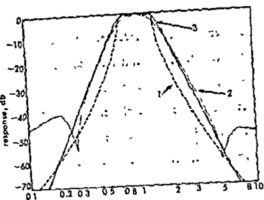


Fig 4 Response characteristics of three octave-band filters: 1, *General Radio* 1550-A, 2, H. H. Scott 420-A; 3, Mine Safety Appliances Sound Scope Abscissa scale is in units of f/f_0, where f is the frequency and f_0 is the center frequency of the octave band (Data furnished by G W Kamperman)

octave frequency ranges or submultiples thereof. Commercially available sound analyzers permit analysis in bands as narrow as ⅒ octave and as wide as 2 octaves However, bandwidths of ⅓, ½, and 1 octave are most common Such sound analyzers almost always employ sets of passive filters, one of which may be selected for use by a switch Continuous tuning of the filter through the frequency range is uncommon, although some narrow constant percentage-bandwidth filters are on the market

The most familiar type of sound analyzer is the octave-band analyzer (OBA), the operation of which has been standardized by the American Standards Association This instrument contains an amplifier, a set of filters selected by a switch, and a meter which can indicate the signal strength in each filter pass band The standard instrument has the following filter bands low pass to 75 cps, 75–150 cps 150–300 cps, 300–600 cps, 600–1200 cps 1200–2400 cps 2400–4800 cps, 4800 cps to high pass and an over-all band which passes all frequencies from 20 to 10,000 cps (see Fig 4) The lower and upper cutoff points are controlled by the amplifier bandwidth, the microphone response, or both Assuming normal values for these limits, the first of these bands is sometimes called the 20–75 cps band and the next-to-last the 4800–10 000 cps band It can be seen that six of these bands span an octave in frequency hence the name However, some of the commercially available versions of this instrument also permit an analysis in narrower frequency bands

The octave-band analyzer is basically a field instrument It is portable and battery operated, and is commonly used to analyze the electrical output of a sound level meter This output is fed either to the over all position or to one of the filters as selected by a switch, and thence to a 10-db step attenuator The output of the attenuator goes to a feedback stabilized amplifier stage whose gain may be set during calibration. Finally, a meter and an output jac e fed through parallel output amplifiers. It is usually possible to vary the damping in the meter with a switch marked fast and slow.

The standard requires that the insertion loss of each octave filter be 3 db or less, that the filter skirts fall off at approximately 20 db/octave or more, and that the maximum rejection be at least 45 db.

Variable-bandwidth filter. This device, a familiar laboratory instrument, is occasionally used as a sound analyzer. The filter permits the independent selection of upper and lower cutoff frequencies so that almost any bandwidth may be obtained.

The variable-bandwidth filter usually consists of several stages of *RC* filters, each separated by buffer amplifiers High-pass and low-pass characteristics are combined to obtain bandpass characteristics, and the *R* and *C* values may be varied to tune the filter.

Although some variable-bandwidth filters of the *RC* type contain peaking circuits to sharpen the filter edges, these filters usually do not have cutoff rates greater than 18–24 db/octave. Thus they may not be suitable for the analysis of sounds having steeply sloped spectra nor for use as very narrow filters See FILTER, ELECTRIC. [B N K]

Bibliography · L. L. Beranek, *Acoustic Measurements*, 1949, C M. Harris (ed), *Handbook of Noise Control*, 1957; A. P G Peterson and L. L. Beranek, *Handbook of Noise Measurement*, 3d ed., 1956

Nomograph

A graphical relationship between a set of variables that are related by a mathematical equation or law The fundamental principle involved in the construction of a nomographic or alignment chart consists of representing an equation containing three variables, $f(u,v,w) = 0$, by means of three scales

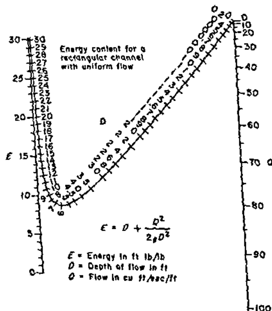


Fig 1 Nomograph for energy content of a rectangular channel with uniform flow

and battery-operated. It is reliable, portable, reasonably stable under battery operation, and light in weight The input impedance is sufficiently high to provide minimal loading for high-impedance microphones. The output impedance is low and the output level high enough to drive most analyzers.

Weighting networks. The weighting networks, of which there are three, adjust the response of the instrument to fall within the limits specified by the American Standards Association. The *A* scale permits the instrument to have a response approximating the 40-phon equal-loudness contour, the *B* scale the 70-phon contour, and the *C* scale provides a flat response to 8000 cps (*see* LOUDNESS; PHON). The *A* and *B* weighting networks are usually used only to make sound-level measurements. In addition to the weighting networks, some instruments have a 20-kc scale which provides a nominally uniform amplifier response up to 20,000 cps. When making a sound-pressure-level measurement or using the sound-level meter as an amplifier to drive a sound analyzer, the *C* scale or the 20-kc scale is used.

The present sound-level-meter standard permits a certain range, or deviation, about the prescribed frequency response curve for each of the weighting networks This varies from more than ±5 db at the extremes of the spectrum to ±2 db in the center. This range is primarily to accommodate variations in the microphones supplied with the instruments.

Meter response. The indicating instrument on a sound-level meter is provided with fast and slow response speeds which may be selected by a damping switch The fast response is standardized so that the meter gives a true indication within about 0.2–0 25 sec after a steady 1000-cps tone is applied to the input of the instrument, and does not overshoot more than 1 db The slow response has not been standardized and may vary between instruments of different manufacture. The indicating instrument is designed to read approximately the rms value of the signal

Output Most sound-level meters have an electrical output as well as an indicating instrument. This permits the use of earphones or some type of sound-analyzing equipment with the SLM The electrical output is either driven from a separate output amplifier or designed to disconnect the indicating instrument when an output cable is connected. This prevents distortion of the output by the nonlinear loading of the indicating instrument

Smaller instruments. Several manufacturers market, under various names, a relatively inexpensive pocket-sized instrument similar to the sound-level meter. Complete with an integral crystal microphone, transistor amplifier, and batteries, this instrument is convenient for initial examinations during noise studies and for acoustical comparisons of different noise sources. However, this device is less accurate than the standard sound-level meter. Likewise, the microphone is not removable, and field calibration of the device is difficult.

Sound analyzer. This is a device which permits the determination of certain properties of a sound.

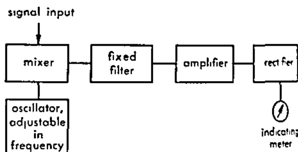


Fig 3. Simplified block diagram of heterodyne analyzer.

In particular, the term sound analyzer is applied to an instrument which permits determination of the amount of sound energy in various frequency bands. This device generally consists of a set of fixed electrical filters, or a tunable electrical filter, along with associated amplifiers and a meter which indicates the filter output. Such a sound analyzer is designed to analyze the electrical signal produced by a microphone in a sound field

Analyzers used for analysis over a range of frequencies are usually divided into three categories: constant bandwidth, constant percentage bandwidth, and variable bandwidth.

Constant-bandwidth analyzers The tunable constant-bandwidth analyzer has a fixed pass band, as the name implies. This fixed pass band is swept through the frequency range of interest, and the electrical signal which passes through the filter at each frequency is observed

A familiar type of constant-bandwidth analyzer is the heterodyne analyzer (Fig 3). In this device, the electrical signal from the microphone modulates (beats with) the signal from an oscillator One of the side bands produced by this modulation is then passed through a fixed filter and detected As the oscillator is tuned, the heterodyne passed by the filter contains components from different frequencies through the range of the microphone signal The heterodyne filter has the advantage of employing a fixed passive filter circuit which may have very sharp cutoff characteristics, for only the oscillator need be tuned to vary the frequency passed by the filter Furthermore, this technique is easily adaptable to use on a linear frequency scale

Narrow, tunable, constant-bandwidth filters are occasionally used as sound analyzers in order to locate discrete frequency components in the sound However, they are much more commonly employed as vibration analyzers, for vibration is usually analyzed over a narrower frequency range than sound The constant-bandwidth filter analyzes the range from 9900 to 10,000 cps in as much detail as the range from 100 to 200 cps. Thus it is not readily adaptable for use on the logarithmic frequency scale common to acoustics. *See* VIBRATION.

Constant-percentage bandwidth filters These have a bandwidth directly proportional to the frequency to which they are tuned As a result, the width of the pass band is constant on a logarithmic frequency scale. Sound an commo stant-percentage

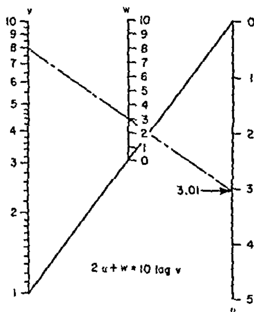


Fig 5 Variation of alignment chart involving logarithm

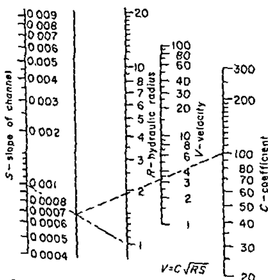


Fig 6 Nomograph for velocity of flow of water in open channels

are constructed in this manner Log-log, semilog and reciprocal coordinate papers are probably the most common [illegible]
binations a [illegible]
graphic ch[illegible]
ables into [illegible] ceeds three.

Types of nomographic charts. The form of the equation serves to classify the type of chart An equation of the form

$$f_1(u) + f_2(v) = f_3(w) \qquad (1)$$

$$f_1(u) + f_2(v) + \cdots = f_n(t) \qquad (1a)$$

leads to a chart of the type shown in Fig 4. The equations of the three scales are $x = m_1 f_1(u)$, $m_2 f_2(v)$ and $[m_1 m_2/(m_1 + m_2)] f_3(w)$, respectively; and $d_1/d_2 = m_1/m_2$ The equation $2u + w = 10 \log v$ is in this form. Taking $m_1 = m_2 = 1$ and $d_1 = d_2$ the scales are $10 \log v$, $-2u$, and $w/2$. The completed chart is shown in Fig. 5. As an example, if $v = 8$ and $w = 3$ the value of u is found to be 3 01.

Another form of equation that leads to a similar chart is

$$f_1(u) f_2(v) \cdots = f_n(t) \qquad (1b)$$

Using logarithms, this equation takes the form of (1) which is

$$\log f_1(u) + \log f_2(v) + \cdots = \log f_n(t) \qquad (1c)$$

The Chezy formula for the velocity of the flow of water $v = c(RS)^{1/2}$ is of this type. When logarithms are used, the equation becomes $\log v = \log c + \frac{1}{2} \log R + \frac{1}{2} \log S$ which may be written

$$\tfrac{1}{2} \log S + \tfrac{1}{2} \log R = \log v - \log c = Q$$

where Q is a dummy variable. The completed chart is shown in Fig 6. To find the value of v when $R = 1$, $S = 0.001$ and $c = 100$, set the straightedge on $S = 0.001$ and $R = 1$; now connect the point of intersection on the dummy scale to $c = 100$ and read $v = 3$ at the point of crossing on the v scale.

Alternatively the equation may be of the form

$$[f_1(u)]^{f_2(v)} = f_3(w) \qquad (1d)$$

Using logarithms, this equation also takes the form of (1), which is

$$\log f_2(v) + \log \log f_1(u) = \log \log f_3(w)$$

A second form is

$$f_1(u) + f_2(v) = f_3(w)/f_4(t) \qquad (2)$$

where $m_1/K = m_3/m_4$ Such a chart is shown in Fig 7

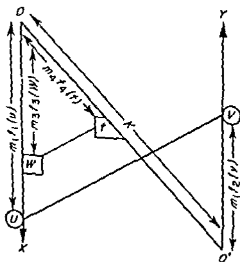


Fig 7 Alignment chart for summation involving a ratio

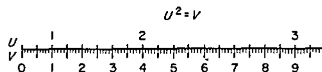


Fig. 2. Conversion scales for finding squares and square roots.

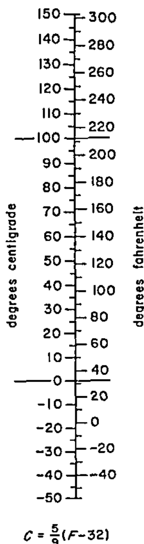


Fig. 3 Stationary scales relate Centigrade and Fahrenheit temperatures.

in such a manner that a straight line cuts the three scales in values of u, v, and w, satisfying the equation. The cutting line is called the isopleth or index line. Numbers may be quickly and easily read from the scales of such a chart even by one unfamiliar with the construction of the chart and the equation involved. Figure 1 illustrates such an example. Assume that it is desired to find the value of E when $D = 2$ and $Q = 50$. Lay a straightedge through 50 on the Q scale and through 2 on the D scale and read 11.8 at its intersection with the E scale. As another example, it might be desired to know what value or values of D should be used if E and Q are required to be 10 and 60 respectively. A straightedge through $E = 10$ and $Q = 60$ cuts the D scale in two points, $D = 2.8$ and $D = 9.4$. This is equivalent to finding two positive roots of the cubic equation $D^3 - 10D^2 + 56.25 = 0$. It is assumed that $g = 32$ ft/sec² in this equation.

Scale. The graphical scale is a curve or straight line, called an axis, on which is marked a series of points or strokes corresponding to a set of numbers arranged in order of magnitude. If the distances between successive strokes are equal, the scale is uniform; otherwise the scale is nonuniform. The scale on a yardstick or thermometer is uniform whereas a logarithmic scale on a slide rule is nonuniform.

Representation of a function by a scale. Consider the function $f(u)$. Lay off, from a fixed point O on a straight line or curve, lengths equal to $f(u)$ units; mark at the strokes indicating the end of each unit the corresponding value of u. If the unit of measure is an inch, the equation of the scale is $x = f(u)$ in. More generally, the equation of the scale is $x = mf(u)$ units, where the constant m (modulus) regulates the length of scale used for the required values of the variable u needed.

Stationary scales. A relation between two variables of the form $v = f(u)$, or $f(v) = F(u)$, can be represented by the natural scales $x = mv$ and $x = mf(u)$; or $x = mf(v)$ and $x = mF(u)$, on the opposite side of the same line or axis. Figure 2 shows the relation $u^2 = v$ using the natural scales $x = mu^2$ and $x = mv$, where in this illustration $m = 0.43$ and the unit is an inch. By using logarithms the above equation becomes $2 \log u = \log v$; and the scales are $x = m(2 \log u)$, and $x = m \log v$.

Adjacent stationary scales may be used to advantage in representing the relationship between the two variables in a conversion formula. Figure 3 shows the relation between degrees Centigrade and Fahrenheit. It is easy to see that $F = 140$ when $C = 60$, and when $F = -40$, $C = -40$.

Perpendicular scales. A relation between two variables u and v of the form $v = f(u)$, $f(u,v) = 0$, or $f(u) = F(v)$ can be represented by constructing two scales $x = mf(u)$ and $y = mF(v)$ on perpendicular axes. Any pair of values of u and v will determine a point in the plane. The locus of all such points is a curve which represents the relationship between the variables u and v. The various types of coordinate paper, sold commercially,

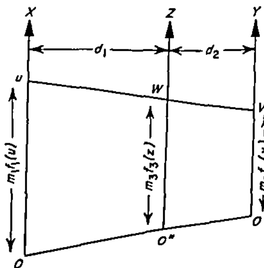


Fig 4. Alignment chart for atio

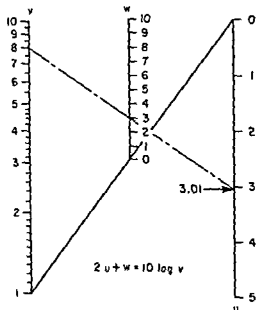


Fig 5 Variation of alignment chart involving logarithm.

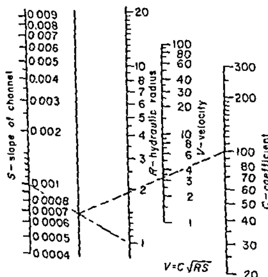


Fig 6 Nomograph for velocity of flow of water in open channels.

are constructed in this manner Log-log, semilog and reciprocal coordinate papers are probably the most common These ...
bina ...
grap ...
able ...
... extends three

Types of nomographic charts. The form of the equation serves to classify the type of chart. An equation of the form

$$f_1(u) + f_2(v) = f_3(w) \qquad (1)$$
$$f_1(u) + f_2(v) + \cdots = f_n(t) \qquad (1a)$$

leads to a chart of the type shown in Fig. 4 The equations of the three scales are $x = m_1 f_1(u)$, ... $m_2/(m_1 + m_2)]f_3(w)$, respectively; and $d_1/d_2 = m_1/m_2$ The equation $2u + w = 10 \log v$ is in this form. Taking $m_1 = m_2 = 1$ and $d_1 = d_2$ the scales are $10 \log v$, $-2u$, and $w/2$. The completed chart is shown in Fig 5. As an example, if $v = 8$ and $w = 3$ the value of u is found to be 3 01.

Another form of equation that leads to a similar chart is

$$f_1(u) f_2(v) \cdots = f_n(t) \qquad (1b)$$

Using logarithms, this equation takes the form of (1) which is

$$\log f_1(u) + \log f_2(v) + \cdots = \log f_n(t) \qquad (1c)$$

The Chezy formula for the velocity of the flow of water $v = c(RS)^{1/2}$ is of this type. When logarithms are used, the equation becomes $\log v = \log c + ½ \log R + ½ \log S$ which may be written

$$½ \log S + ½ \log R = \log v - \log c = Q$$

where Q is a dummy variable. The completed chart is shown in Fig 6 To find the value of v when $R = 1$, $S = 0.001$ and $c = 100$, set the straightedge on $S = 0\,001$ and $R = 1$; now connect the point of intersection on the dummy scale to $c = 100$ and read $v = 3$ at the point of crossing on the v scale.

Alternatively the equation may be of the form

$$[f_1(u)]^{f_2(t)} = f_3(w) \qquad (1d)$$

Using logarithms, this equation also takes the form of (1), which is

$$\log f_2(t) + \log \log f_1(u) = \log \log f_3(w)$$

A second form is

$$f_1(u) + f_2(v) = f_3(w)/f_4(t) \qquad (2)$$

where $m_1/K = m_3/m_4$. Such a chart is shown in Fig 7

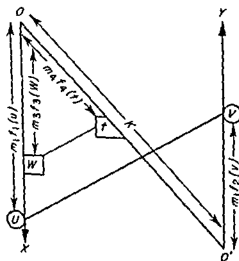


Fig 7. Alignment chart for summation involving a ratio.

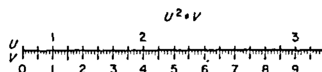


Fig. 2. Conversion scales for finding squares and square roots.

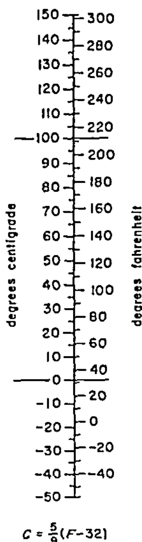


Fig 3 Stationary scales relate Centigrade and Fahrenheit temperatures

in such a manner that a straight line cuts the three scales in values of u, v, and w, satisfying the equation. The cutting line is called the isopleth or index line. Numbers may be quickly and easily read from the scales of such a chart even by one unfamiliar with the construction of the chart and the equation involved Figure 1 illustrates such an example. Assume that it is desired to find the value of E when $D = 2$ and $Q = 50$. Lay a straightedge through 50 on the Q scale and through 2 on the D scale and read 11.8 at its intersection with the E scale As another example, it might be desired to know what value or values of D should be used if E and Q are required to be 10 and 60 respectively. A straightedge through $E = 10$ and $Q = 60$ cuts the D scale in two points, $D = 2.8$ and $D = 9.4$. This is equivalent to finding two positive roots of the cubic equation $D^3 - 10D^2 + 56.25 = 0$. It is assumed that $g = 32$ ft/sec² in this equation.

Scale. The graphical scale is a curve or straight line, called an axis, on which is marked a series of points or strokes corresponding to a set of numbers arranged in order of magnitude. If the distances between successive strokes are equal, the scale is uniform; otherwise the scale is nonuniform. The scale on a yardstick or thermometer is uniform whereas a logarithmic scale on a slide rule is nonuniform.

Representation of a function by a scale Consider the function $f(u)$. Lay off, from a fixed point O on a straight line or curve, lengths equal to $f(u)$ units, mark at the strokes indicating the end of each unit the corresponding value of u. If the unit of measure is an inch, the equation of the scale is $x = f(u)$ in. More generally, the equation of the scale is $x = mf(u)$ units, where the constant m (modulus) regulates the length of scale used for the required values of the variable u needed.

Stationary scales. A relation between two variables of the form $v = f(u)$, or $f(v) = F(u)$, can be represented by the natural scales $x = mv$ and $x = mf(u)$; or $x = mf(v)$ and $x = mF(u)$, on the opposite side of the same line or axis. Figure 2 shows the relation $u^2 = v$ using the natural scales $x = mu^2$ and $x = mv$, where in this illustration $m = 0.43$ and the unit is an inch. By using logarithms the above equation becomes $2 \log u = \log v$, and the scales are $x = m(2 \log u)$, and $x = m \log v$.

Adjacent stationary scales may be used to advantage in representing the relationship between the two variables in a conversion formula. Figure 3 shows the relation between degrees Centigrade and Fahrenheit. It is easy to see that $F = 140$ when $C = 60$; and when $F = -40$, $C = -40$

Perpendicular scales A relation between two variables u and v of the form $v = f(u)$, $f(u,v) = 0$, or $f(u) = F(v)$ can be represented by constructing two scales $x = mf(u)$ and $y = mF(v)$ on perpendicular axes Any pair of values of u and v will determine a point in the plane The locus of all such points is a curve which represents the relationship between the variables u and v The various types of coordinate paper, sold commercially,

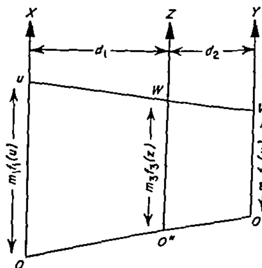


Fig. 4. Alignment chart for [illegible] atio

$$\begin{vmatrix} f_1(u) & g_1(u) & 1 \\ f_2(t) & g_2(t) & 1 \\ f_3(w) & g_3(w) & 1 \end{vmatrix} = 0$$

then a nomograph can be constructed from the parametric equations

$$x = f_1(u),\ y = g_1(u)$$
$$x = f_2(t),\ y = g_2(t)$$
$$x = f_3(w),\ y = g_3(w)$$

The mean temperature equation

$$T = (T_1 - T_2)/(\ln T_1 - \ln T_2)$$

expressed as a determinant is

$$\begin{vmatrix} 0 & T & 1 \\ 1/\ln T_1 & T_1/\ln T_1 & 1 \\ 1/\ln T_2 & T_2/\ln T_2 & 1 \end{vmatrix} = 0$$

Figure 11 shows the resulting nomograph.

Figure 1 could also have been constructed from the determinant form after the energy equation is written as

$$\begin{vmatrix} 0 & (12.2/100^2)Q^2 & 1 \\ 9 & -E/3 & 1 \\ \dfrac{9D^2}{D^2 + \dfrac{100^2}{73.2g}} & \dfrac{-D^3}{3D^2 + \dfrac{100^2}{73.2g}} & 1 \end{vmatrix} = 0$$

Circular nomographs. The general form of the basic determinant for a circular nomograph is

$$\begin{vmatrix} \dfrac{1}{1 + [f_1(u)]^2} & \dfrac{f_1(u)}{1 + [f_1(u)]^2} & 1 \\ \dfrac{1}{1 + [f_2(t)]^2} & \dfrac{f_2(t)}{1 + [f_2(t)]^2} & 1 \\ \dfrac{1}{1 + f_3(w)} & 0 & 1 \end{vmatrix} = 0$$

where the u and t scales lie on circular axes having the same center and with radii equal to ½. Consider the equation $AB \cos(\theta - \phi) + BU$ $(\sin\phi - \cos\theta) + AU (\sin\theta - \cos\phi) = 0$. Expressed in the form of a determinant, it is

$$\begin{vmatrix} A \sin\theta & A\cos\theta & 1 \\ B\cos\phi & B\sin\phi & 1 \\ U & U & 1 \end{vmatrix} = 0$$

which leads to the nomograph of Fig. 12. [R.D.D.]

Bibliography: W. Staniar, *Plant Engineering Handbook*, 1950; R. D. Douglass and D. P. Adams, *Elements of Nomography*, 1947.

$$T = \frac{T_1 - T_2}{\ln T_1 - \ln T_2}$$

Fig. 11 To find mean temperature, lay straightedge between two given temperatures on curved scale, read mean temperature on vertical scale.

example

$\theta = 30°$

$\phi = 315°$

cut U at 2.1

Fig. 12 Circular nomograph results from trigonometric equation.

Nonelectrolyte

A chemical compound which does not conduct electricity either when fused or when dissolved in a solvent such as water. The bonding in nonelectrolytes is covalent. Most organic compounds, with the exception of the acids and amines, are nonelectrolytes, and so are many inorganic compounds such

[T.C.W.]

Nonmetal

The elements are conveniently, but arbitrarily, divided into metals and nonmetals. The nonmetals do not conduct electricity readily, are not ductile

A third form is

$$f_1(u) + f_2(v)f_3(u) = f_4(u) \qquad (3)$$

where $$x_1 = \frac{m_1K}{m_1f_3(u) + m_2}f_3(u)$$

and $$y_1 = \frac{m_1m_2}{m_1f_3(u) + m_2}f_4(u)$$

Such charts are illustrated by Figs. 1 and 8. This is a frequently encountered type of equation. In constructing the nomograph of Fig. 1, first the equation $E = D + Q^2/2gD^2$ was rewritten in the form $-Q^2 + 2gD^2E = 2gD^3$ which corresponds to the basic form of (3).

A fourth basic form is

$$f_1(u)/f_2(v) = f_3(w)/f_4(t) \qquad (4)$$

where $m_1/m_2 = m_3/m_4$. This can be treated as a Z-chart by using natural scales, or it takes the form of (1) by using logarithms.

A fifth form is the conversion chart

$$f_1(u) = Cf_2(v) \qquad (5)$$

This is a special case of (1) or (4) and is useful when several conversions are to be made. Examples are Figs. 9 and 10. In practice it may be useful or even necessary to use two or more combinations of these basic types.

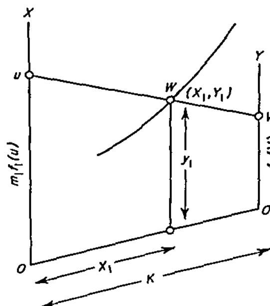


Fig. 8. Alignment chart for summation involving a product.

Determinant as a basis of a nomograph. The condition that the three points (x_1,y_1), (x_2,y_2) and (x_3,y_3) lie on a straight line is that the determinant

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0$$

If an equation that relates three variables u, v, and w can be expressed in the determinant form

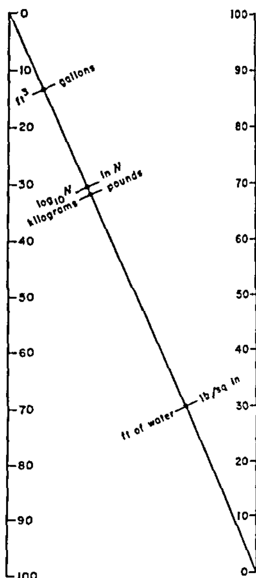


Fig. 9. Conversion chart relates units of measure on linear scales.

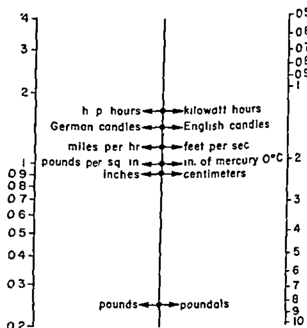


Fig. 10. Conversion chart relates units of measure on exponential scale.

$$\begin{vmatrix} f_1(u) & g_1(u) & 1 \\ f_2(t) & g_2(t) & 1 \\ f_3(u) & g_3(u) & 1 \end{vmatrix} = 0$$

then a nomograph can be constructed from the parametric equations

$$x = f_1(u),\ y = g_1(u)$$
$$x = f_2(t),\ y = g_2(t)$$
$$x = f_3(u),\ y = g_3(w)$$

The mean temperature equation

$$T = (T_1 - T_2)/(\ln T_1 - \ln T_2)$$

expressed as a determinant is

$$\begin{vmatrix} 0 & T & 1 \\ 1/\ln T_1 & T_1/\ln T_1 & 1 \\ 1/\ln T_2 & T_2/\ln T_2 & 1 \end{vmatrix} = 0$$

Figure 11 shows the resulting nomograph

Figure 1 could also have been constructed from the determinant form after the energy equation is written as

$$\begin{vmatrix} 0 & (12.2/100^2)Q^2 & 1 \\ 9 & -E/3 & 1 \\ \dfrac{9D^2}{D^2 + \dfrac{100^2}{73.2g}} & \dfrac{-D^2}{3D^2 + \dfrac{100^2}{73.2g}} & 1 \end{vmatrix} = 0$$

Circular nomographs. The general form of the basic determinant for a circular nomograph is

$$\begin{vmatrix} \dfrac{1}{1 + [f_1(u)]^2} & \dfrac{f_1(u)}{1 + [f_1(u)]^2} & 1 \\ \dfrac{1}{1 + [f_2(t)]^2} & \dfrac{f_2(t)}{1 + [f_2(t)]^2} & 1 \\ \dfrac{1}{1 + f_3(u)} & 0 & 1 \end{vmatrix} = 0$$

where the u and v scales lie on circular axes having the same center and with radii equal to ½. Consider the equation $AB \cos(\theta - \phi) + BU$

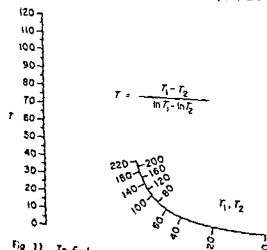
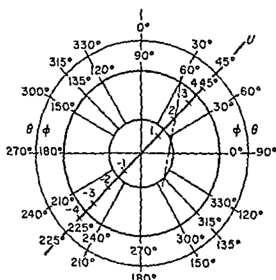


Fig 11 To find mean temperature, lay straightedge between two given temperatures on curved scale, read mean temperature on vertical scale.



example
$\theta = 30°$
$\phi = 315°$
cut U at 2.1

Fig 12. Circular nomograph results from trigonometric equation

$(\sin\phi - \cos\theta) + AU \ (\sin\theta - \cos\phi) = 0$. Expressed in the form of a determinant, it is

$$\begin{vmatrix} A\sin\theta & A\cos\theta & 1 \\ B\cos\phi & B\sin\phi & 1 \\ U & U & 1 \end{vmatrix} = 0$$

which leads to the nom[illegible]

Nonelectrolyte

A chemical compound which does not conduct electricity either when fused or when dissolved in a solvent such as water The bonding in nonelectrolytes is covalent. Most organic compounds, with the exception of the acids and amines, are nonelectrolytes, and so are many inorganic compounds such as the halides of the nonmetal[illegible]

[T.C.W.]

Nonmetal

The elements are conveniently, but arbitrarily, divided into metals and nonmetals. The nonmetals do not conduct electricity readily, are not ductile, do not have a complex refractive index, and in general have high ionization potentials

The nonmetals may vary widely in physical properties. Hydrogen is a colorless permanent gas; bromine is a dark red, volatile liquid, and carbon, as diamond, is a solid of great hardness and high refractive index If the periodic table is divided

A third form is

$$f_1(u) + f_2(t)f_3(w) = f_4(w) \qquad (3)$$

where $x_1 = \dfrac{m_1 K}{m_1 f_3(w) + m_2} f_3(w)$

and $y_1 = \dfrac{m_1 m_2}{m_1 f_3(w) + m_2} f_4(w)$

Such charts are illustrated by Figs. 1 and 8. This is a frequently encountered type of equation. In constructing the nomograph of Fig. 1, first the equation $E = D + Q^2/2gD^2$ was rewritten in the form $-Q^2 + 2gD^2E = 2gD^3$ which corresponds to the basic form of (3).

A fourth basic form is

$$f_1(u)/f_2(v) = f_3(w)/f_4(t) \qquad (4)$$

where $m_1/m_2 = m_3/m_4$. This can be treated as a Z-chart by using natural scales, or it takes the form of (1) by using logarithms.

A fifth form is the conversion chart

$$f_1(u) = Cf_2(v) \qquad (5)$$

This is a special case of (1) or (4) and is useful when several conversions are to be made. Examples are Figs. 9 and 10. In practice it may be useful or even necessary to use two or more combinations of these basic types.

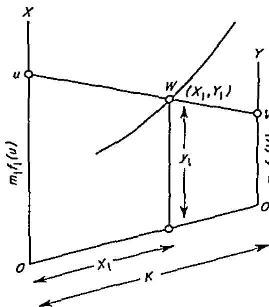


Fig. 8. Alignment chart for summation involving a product

Determinant as a basis of a nomograph. The condition that the three points (x_1,y_1), (x_2,y_2) and (x_3,y_3) lie on a straight line is that the determinant

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0$$

If an equation that relates three variables u, v, and w can be expressed in the determinant form

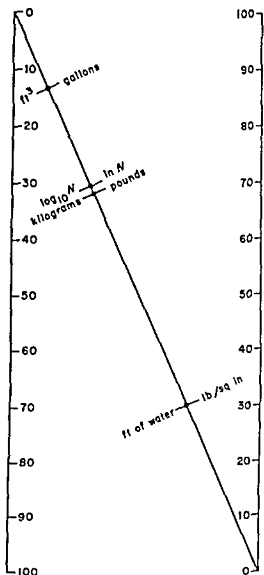


Fig 9. Conversion chart relates units of measure on linear scales

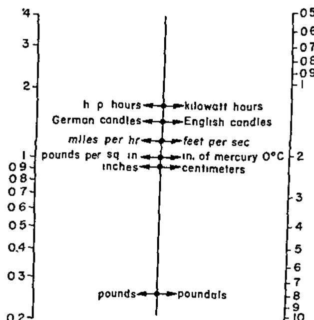


Fig 10 Conversion chart relates units of measure on exponential scale



Fig. 1 Landform map of North America (Drawn by E. Raisz)

diagonally from upper left to lower right, all the nonmetals are on the right-hand side of the diagonal. Examples of elements which do not fit neatly into this useful but arbitrary classification are tin, which exists in two allotropic modifications, one definitely metallic and the other with many properties of a nonmetal, and tellurium and antimony. Such elements are called metalloids. *See* IONIZATION POTENTIAL; METAL; METALLOID; PERIODIC TABLE. [T.C.W.]

Nonstoichiometric compounds

The law of constant proportions is usually considered to be among the most general principles in chemistry. It is, of course, true that the familiar organic compounds and many inorganic substances appear to have constant compositions. There are, *however, a large number of inorganic substances* which do not exhibit constant, integral ratios of atoms in their stoichiometries. Many familiar binary compounds exhibit this variability. Examples are $VH_{0.56}$, $CeH_{2.69}$, $TiO_{0.6-1.35}$, $FeO_{1.06-1.19}$, and $Cu_2S_{1.18}$. These substances are called nonstoichiometric or Berthollide compounds. The classification does not normally include simple solid solutions such as are common to certain alloy systems, because these are not compounds in the usual meaning of the word. *See* DEFINITE COMPOSITION, LAW OF; STOICHIOMETRY.

Classification. The deviations from stoichiometry may be explained and classified simultaneously. The most obvious classification of these compounds involves separating them into two categories depending on the nature of the element which is in excess, that is, those compounds containing an excess of the more electropositive element (metal) and those containing an excess of the more electronegative element (nonmetal). Because these deviations from ideal stoichiometry are limited to the solid state, they are most readily understood in terms of the manner in which the structures of crystals may be altered by the presence of an excess of one of the elements. These effects are understandable as a consequence of theories dealing with defects in crystalline solids. The two types of crystal defects which are of significance here involve the removal of ions from their usual sites in the crystal, giving rise to ion vacancies, and the addition of ions into voids (interstices) between the ions that make up a normal completed lattice. The latter may be referred to as interstitial ions. These considerations double the number of classes of nonstoichiometric compounds, as follows:

Compounds with excess metal: Type I, as a result of vacancies at anion sites; type II, as a result of the presence of interstitial cations.

Compounds with excess nonmetal: Type III, as a result of vacancies at cation sites; type IV, as a result of the presence of *interstitial anions*. This classification is illustrated in the table.

Structure. In all of these cases, crystals must remain electrostatically uncharged. Consequently, an excess of metal can occur only if some of the metal ions are in lower oxidation states than the bulk of those making up the lattice, or alternatively, if the number of electrons necessary to maintain electrical neutrality are otherwise incorporated in the lattice. It has been shown that in some type I structures, electrons may occupy the anion vacancies. Similarly, if the nonmetal is in excess, an equivalent number of the metal ions must exist in oxidation states higher than the dominant state.

From these considerations, it may be concluded that nonstoichiometric compounds should occur when (1) the energy necessary to produce crystal defects is small, (2) the energy difference between two oxidation states of one of the elements is not too large, and (3) the difference in the sizes of the atoms is similar for the two oxidation states mentioned in (2). *See* CRYSTAL DEFECTS; IONIC CRYSTALS, SOLID-STATE CHEMISTRY. [D H B]

Classes of nonstoichiometric compounds

Compounds with excess metal							
Type I				Type II			
M^+	X^-	M^+	X^-	M^+	X^-	M^+	X^-
X^-	M^+	X^-	M^+	X^-	M^+	X^-	M^+
					M^+		
M^+		M^+	X^-	M^+	X^-	M^+	X^-
X^-	M^+	X^-	M^+	X^-	M^+	X^-	M^+
Example: KCl, TiO				Example: ZnO, CdO			
Compounds with excess nonmetal							
Type III				Type IV			
M^+	X^-	M^+	X^-	M^+	X^-	M^+	X^-
X^-		X^-	M^+	X^-	M^+	X^-	M^+
					X^-		
M^+	X^-	M^{2+}	X^-	M^+	X^-	M^{2+}	X^-
X^-	M^+	X^-	M^+	X^-	M^+	X^-	M^+
Example: Cu_2O, FeS				Example: UO_2			

M^+, unipositive cation, X^-, uninegative anion; open space, vacancy

Normal

A term generically synonymous with perpendicular, which often refers, specifically, to a line that goes through a point P of a curve C and is perpendicular to the tangent to C at P. If a plane curve C has equation $y = f(x)$, in rectangular coordinates, the normal (line) to C at $P(x_0,y_0)$ has slope $-1/f'(x_0)$, provided $f'(x_0) \neq 0$. The expression $f'(x_0)$ denotes the derivative of $f(x)$, evaluated for $x = x_0$, and so has equation $y - y_0 = [-1/f'(x_0)](x - x_0)$. If curve C is not a plane curve, all normal lines of C at point P on C lie in a plane, the normal plane of C at P. For the other uses of normal (for example, normal form of equation of a line or a plane), *see* ANALYTIC GEOMETRY. [L.M.BL.]

North America

The northern of the two continents of the New World or Western Hemispher tending fro narrow parts in the tropics t y b

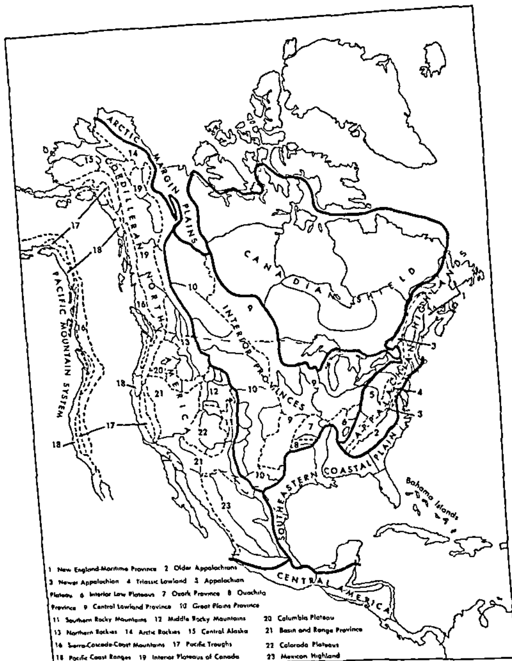


Fig 2 Sketch map of physical divisions of North America *(Modified from A K Lobeck and others)*

sion over long periods has resulted in a surface that has been somewhat smoothed and lowered in relief so it resembles an undulating plain, except along the southern and eastern borders. Most of it lies at elevations of less than 2000 ft above sea level. Near the coast of Labrador uplift and tilting have created higher elevations which make the area appear mountainous. The escarpment along the Atlantic Coast has been serrated by the tongues of mountain glaciers into fiords, making this area the most rugged section of the province. See FIORD

Two centers of Pleistocene glaciation were located in the Ungava Peninsula (between the Atlantic Ocean and Hudson Bay), known as the Labrador center, and the Keewatin center, west of Hudson Bay. The glaciers scraped most of the loose rock off the Laurentian Upland, leaving outcrops of bare rock on knobs and low rises and gouging out troughs and depressions in the softer

and short cool summers, bacterial action is slow so that peat and muck bogs are readily formed. Th[e] great northern or coniferous forest is well devel[-] oped in the southern portions of the Laurentia[n] Upland and in areas of better soils, but the tree[s] become more scattered and stunted in growth wi[th]

portions in middle latitudes and Arctic polar margins. This protuberance of the earth's crustal shell stands with some 9,363,000 mi² of land above the sea and ranks third in size among the continents. The continental mass upstands with deep submarine continental scarps toward the ocean basins at a marginal break or change in slope which appears at varying shallow depths of continental submarine shelf. These shelf areas of shallow sea bottom extend offshore for distances up to several hundred miles (*see* CONTINENT). As a practical expedient, maps commonly show the bathymetric contour of 500 ft or 100 fathoms (600 ft) to indicate the approximate limits of these peripheral continental features (Fig 1). Most of this outline of North American physical geography emphasizes physiography, especially land surface character, but it also includes some aspects of continental shelves considered as coastal-zone character. This focus is intended to present a basis for understanding other environmental and human features with which the land character is commonly interrelated.

Location. Two aspects of the location of North America, local and global, appear of considerable importance in present affairs.

Local extent and location. North America, the continent with adjacent islands, traditionally includes Greenland and the West Indies which are considered only incidentally here. *See* ARCTIC AND SUBARCTIC ISLANDS; WEST INDIES.

The North American continent extends from southeastern Panama (6°6′N) at its isthmian frontier with South America to the northern tip of the Boothia Peninsula (72°10′N), northeast of the magnetic pole. This makes an extreme north-south distance of approximately 4554 miles If the Arctic islands are included, this distance increases to 5272 miles; the northern tip of Grant Land on Ellesmere Island is at 82°30′N. The maximum longitudinal spread is from easternmost Greenland (18°W) to the westernmost of the Aleutian Islands (172°E), a total of 170° Canada and the United States, exclusive of Alaska, lie between longitude 52°W and 142°W Since the narrower portion of the continent lies equatorward of 30°N latitude, relatively little of its area is in the tropics, whereas the greatest area occurs in mid-latitudes with considerable portions extending into the subpolar and polar regions

Global position. The peculiarly strategic position of this northern continent in relation to Central and South America, the oceans, and the other continents of the world is difficult to visualize without the aid of a globe (*see* GLOBE, TERRESTRIAL) Almost all of North America lies west of South America. Aviation routes due south from most of North America would miss South America entirely; those south from New York and Miami cross Panama, continue along the western coastal margins of South America, and turn eastward over the Andes Mountains, in reaching such places as Buenos Aires or Montevideo. The distance is shorter, by surface ship, between the east and west coasts of North America via the Panama Canal, and it is also shorter from Buenos Aires to New Orleans via Cape Horn or the Strait of Magellan, along the west coast, and through the Panama Canal, than via the routes of eastern South America into the Caribbean Sea and Gulf of Mexico The location of Panama and the Caribbean is still strategic with regard to modern airways as well as to sea routes between the Atlantic and Pacific Oceans.

The grouping of the broad parts of the land masses of the Northern Hemisphere, North America and Eurasia, around the polar region and the Arctic Ocean has special significance today. Great circle routes between the regions of greatest population in North America, Europe, and southern and eastern Asia have long caused surface ships to swing far north along the Aleutians, Newfoundland, Greenland, and Iceland on the subarctic margins (*see* GREAT CIRCLE, TERRESTRIAL). Modern aircraft, submarines, and navigation methods make possible today even shorter great circle routes between certain centers via polar-margin and transpolar routes which are more northerly than was previously considered feasible. *See* NAVIGATION, POLAR NAVIGATION; SUBMARINE.

Physiographic provinces. The land surface is the base upon which all the other geographical elements are placed. North America can be reasonably divided into some 6 broad physiographic divisions, based on their general geologic structure and subsequent landform development. These are, in turn, further subdivided on the basis of more localized variations in topography. Various phases of physical geography are integrated into a discussion of the physiographic divisions so as to present some composite geographic concept of each division, province, and section.

CANADIAN SHIELD

The Canadian Shield, also known on the mainland as the Laurentian Upland province, is the core around which the rest of North America has been built. In general, this province occupies the northeast quarter of North America, being defined by the St. Lawrence Gulf and River, the upper Great Lakes, and the series of large lakes in central Canada extending north from the United States border, including Winnipeg, Athabaska, Great Slave, and Great Bear lakes The two extensions of this region into the United States are the Adirondack Mountains in northern New York state and the Superior Highland circling the western end of Lake Superior.

The Canadian Shield is composed of ancient crystalline rocks which have been subjected to geological processes such as volcanism, folding, faulting, warpage, and erosion. With minor exceptions the surface rocks were formed in Archeozoic and Proterozoic times (*see* GEOLOGICAL TIME SCALE). It is here, outcropping at the surface, that some of the most ancient rocks of the world are found, and any younger overlying structures are missing, having been eroded and carried · Extensiv



ig 4 Map of natural vegetation of North America (From P E James and H V B Kline, Jr, Geography of Man, Ginn, 1949)

Ungava Bay in the north to the St. Lawrence Bay To the east lie the higher mountainous portions embracing the Labrador coast. Both coniferous and broad leafed trees grow in the area but the trees generally are not well developed and give place to tundra in the north and to an almost vegetationless zone along the Labrador coast (Fig 3). Communications have been very poor so the area remains practically undeveloped

Keewatin and Barren Lands. The Keewatin section lies west of Hudson Bay, and the Barren Lands are the tundra extending north to the inter-

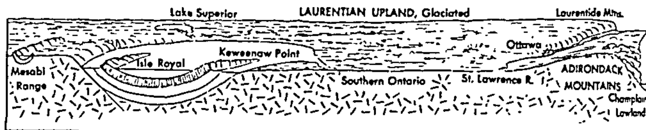


Fig 3. Three-dimensional block profile of surface and structure of a part of the Laurentian or Canadian Shield. (*From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957*)

progression northward. *See* FOREST VEGETATION; TUNDRA.

Native animal life is varied and plentiful with the larger herbivores (vegetation eaters) represented by moose, bears, deer, caribou, and musk oxen, and the carnivores (meat eaters) involving a great variety of smaller fur-bearing animals, such as wolves, fox, otter, and mink. Fish in great variety are found in the many lakes and streams, while numerous migratory land birds are attracted during the summer by the hordes of mosquitoes and insects that annually hatch in the swamps and bogs. Migratory waterfowl find this area ideal for their breeding grounds and summer sustenance.

Because of the low average annual temperatures, evaporation is slight, and a small amount of precipitation is capable of keeping the area humid. Early snowfalls do not melt and a considerable snow depth is accumulated by late winter even when the total winter precipitation represents only 3 or 4 in. of rainfall.

Most of the rivers are of local significance only, having glacial lakes as their origin and containing many rapids and falls. The rapid run-off of melting snow in the spring creates floods, whereas in winter the surfaces may be frozen over. Along the southern border near settled areas some streams are utilized for the generation of hydroelectric power, but most of this resource is as yet undeveloped.

The Canadian Shield is one of the great mineralized areas of the world, containing a variety of both ferrous and nonferrous ores Detailed prospecting and utilization of the minerals are generally limited to the outer border of this province, especially along the south fringe where economical transportation is provided.

Podzol soils have profiles of 18 to 24 in. in the equatorward sections but northward become progressively shallower until the tundra soils are reached. In general, the soil is thin, often coarse in texture, acid in reaction, and of questionable fertility for agriculture. Many areas are waterlogged so that mature profile soils cannot be formed. *See* SOIL; SOIL, ZONAL DISTRIBUTION.

Adirondacks. This detached section in New York state is an eroded plateau on the northwest but is a rugged mountain area with elevations over 5000 ft in the central and southeastern sections. The area has been made a state park, the largest in the United States. Timber cover of mixed hardwoods and conifers is maintained to enhance its value for recreation and to protect the watershed which supplies some of the water needed by New York City. Iron ore is mined in the northern part of this subdivision.

Superior Highland. A subdued portion of the Canadian Shield encircles the western end of Lake Superior. It is best known for the iron-ore deposits which have been extensively exploited in Minnesota, Michigan, and Wisconsin and for native (metallic form) copper, that, in all of North America, is found in commercial quantities only in the Keweenaw Peninsula of Michigan.

Clay Belt. Bordering the western and southern portions of Hudson and James Bays, the Clay Belt is a lowland area composed of recently deposited clays and silts laid down in large glacial lakes during the withdrawal of the continental glaciers. These clays are underlain by sedimentary rocks of Paleozoic age (500,000,000 to 250,000,000 years old) that slope gently under Hudson Bay. This area has only recently been uplifted above sea level West and south from Hudson Bay the area rises gradually to an elevation of a few hundred feet. The low elevation and almost level nature of this area give rise to many swamps. Better soils, of the clays and silts, provide more favorable conditions for forest growth so that some of the best spruce stands in the Canadian Shield are found in this section.

Ontario-Quebec subdivision. This section is best known for its mineral and forest resources The geology is very complex so that great variations occur within limited distances Gold is widespread in distribution but ores of many nonferrous metals have been uncovered at numerous sites Sudbury accounts for 85% of the nickel mined in the world, excluding the Communist countries Ores of copper, lead, zinc, and cobalt, and silver deposits are also worked. The spruce forests are utilized for wood pulp and paper Sportsmen and other recreationalists are attracted by this land and by the game and fish which abound in the forests, lakes, and streams of the area

Labrador section. The Labrador section has been roughly prospected and holds promise of being as well endowed with minerals as the Ontario-Quebec section, but until recently very little has been done to exploit the mineral resources. Iron ores in the Ungava trough are now being min
This trough extends as a do ulted are

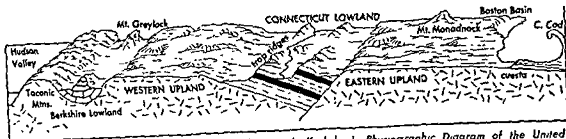


Fig 5 Diagram of land surface and structural relationships in southern part of New England (From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

s more east-west and the rocks are geologically younger. Very little of the area rises above 500 ft in elevation so that local relief is moderate Lowlands between the ridges in Nova Scotia provide better-quality soils which are the sites for agricultural settlements Coal, mined in Cape Breton Island, is the base for a small-scale steel industry Underlying crystalline rocks, exposed along the ocean shoreline by wave cutting, provide a very irregular shore with small bays, headlands, and islands The coastal strip, which becomes narrower to the south, is covered by unconsolidated marine clays which are intensively farmed in New Brunswick and southern Maine but elsewhere have reverted to second-growth scrub forest after lumbering Fishing and boat building support numerous small settlements along the coast A few large cities such as Boston and Portland depend on trade and manufacture of raw materials imported from outside this section Glaciation contributed to the irregularity of the coastline The continental shelf here is the wide Grand Banks noted for fishing and for stormy seas

New England Upland section Old rocks of this area constitute a core of the New England-Maritime province The crystalline rocks have been subject to a number of peneplanations (*see* GEOMORPHOLOGY, FLUVIAL EROSION CYCLE) during and since Paleozoic time so that the area is a plateau with the hill crests varying from 800 to 2000 ft in elevation Beginning on the shore of Baie de Chaleur, the Upland extends south to Long Island Sound with a prong forming Manhattan Island and another protruding southwestward across the Hudson River to the vicinity of Reading, Pennsylvania The southern portions of this section are lower than the rest, ranging from sea level to elevations of a few hundred feet, and creating hilly belts rather than plateaus Forests are maintained in the central and northern sections while efforts to reforest the area are progressing in the south.

Connecticut Lowland A down-faulted and tilted block of Triassic rocks through which the Connecticut River flows makes an area of fertile soil and divides the southern upland into two segments.

Taconic section This is a narrow strip lying between the New England Upland and the Green Mountains to the east and the Hudson Champlain Lowland to the west (Fig 5) It is an area of folded sediments that have in part been metamorphosed creating slate from shales. The higher ridges reach 2000 ft elevation compared to the lower relief in the western Upland.

Green Mountain section. Some of Vermont and southern Quebec resembles the Taconic section in that it is a folded area with the axis running north-south, but the Green Mountain folding was more intensive so that most of the rocks are metamorphosed and intrusions of volcanics provide a greater variety of rocks. More resistant ridges reach 3000 ft elevation. Relatively little lowland is available for agriculture; the main activity is mining granite, marble, and, in Canada, asbestos.

White Mountain section. Here the surface rises above the New England Upland in northern New Hampshire and western Maine to an elevation of 6000 ft It trends in a northeasterly direction from the Green Mountains, but is composed of individual erosional remnants rather than parallel ridges Higher elevations are above timber line and the heavy snows in this section have encouraged its use for winter sports.

Gaspé section This coastal upland resembles the Taconic Mountains except that the folds extend eastward around the north portion of the New England area to the Gulf of St. Lawrence. Here elevations reach 4250 ft. The sparse population is concentrated in small villages along the coast

Appalachian provinces. Northwest and west of the New England-Maritime province are structures of folded Paleozoic rock in the St. Lawrence Val-

istics are such that it is logical to consider them related parts of the northeast-southwest extent of the Newer Appalachians, variously called the folded valley and ridge province, the Great Valley, and numerous other names in local parts (such as Shenandoah Valley). The regionally parallel ridges and valleys are subdued, but present, in these northeastern parts of the Newer Appalachians, but the ridges of relatively resistant rock upstand in linear pattern of hill and low mountain terrain between Pennsylvania and Alabama

A sketch of geological history may be of aid in understanding the Appalachian provinces. Thousands of feet of sediments were scoured off from a great Paleozoic Appalachian highland and deposited into beds to the westward in what is now interior North America. These beds extend great dis-

island waters of the Arctic archipelago. These areas are little-known. The southern part appears to be a poorly drained, low plateau with a thin cover of coniferous forest. In the north, crystalline rock slopes gently northward and forest gives way to tundra. These Barren Lands are the home of caribou and musk oxen. Some uranium is mined in the south, in Manitoba, and on the shores of Great Bear Lake in the north. Adequate prospecting should uncover additional mineral resources.

APPALACHIAN HIGHLANDS

A great variety of highlands and associated lowlands in eastern United States and southeastern Canada are considered as composing the Appalachian Highlands, making it a very complex region. It is generally divided into the New England-Maritime Canada province and the Appalachian province, which in turn is further subdivided into older and newer sections

Rocks ranging from ancient crystalline to Recent sediments are found in the same geographic area in various sections of this highland region. The crystalline rocks predominate in the northern and eastern portions while the northwest and west are characterized by sediments of Paleozoic time. *See* GEOLOGICAL TIME SCALE

Geologic structures of almost every type, which have been subjected to a variety of erosional and depositional processes, are found somewhere in this region Doming, folding, faulting, and tilting, as well as volcanic activities, contribute to the variety of surface features. Major uplifts followed by long periods of erosion have left what might otherwise have been a very high and rugged mountain area as a generally subdued highland with maximum elevation only slightly exceeding 6000 ft in a few summits. Even the peaks are smoothed and rounded by erosional forces. The northeast-southwest trend of the province from Newfoundland to northern Alabama appears in most of the individual structures.

The severe winter phase of the continental humid climate prevails throughout the area providing marked seasonal changes between winter and summer. The more elevated sections in the south have cooler temperatures than the latitudinal position would indicate, thus providing relief from summer heat in adjacent lowlands. Also, the uplands receive from 35 to 55 in. of precipitation which makes this a well-watered region and in winter provides a deep snow blanket

Podzol and gray podzolitic soils prevail where slopes permit the accumulation of regolith (fragmental mineral material) and erosion has been less rapid than the soil-making processes. Where formed from crystallines, these soils are acidic in reaction but are neutral or alkaline in the limestone outcrops which provide the better soils for agriculture.

The mixed mid-latitude broad-leaf coniferous forest occurs throughout the region, but locally either the broad-leaf or coniferous species may occur in pure stands with different species predominating (Fig. 4). Excessive lumbering has changed the character of the remaining forest, even on the more rugged and less settled parts of the region.

The larger animals native to the Appalachians include bear and deer, and a few moose in the north, and considerable numbers of both larger and smaller animals still live in the area. Small game, fur bearers, and meat animals include beaver, fox, opossum, and squirrels.

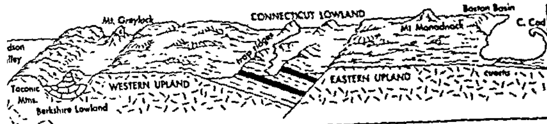
The rivers are mostly small and turbulent. The relatively heavy rainfall supplies abundant water that drains out of the region and the upland character of the area provides many small hydroelectric-power sites. The water resources, however, have been useful in attracting industries that utilize large amounts of water, such as papermaking and the synthetic fiber industry.

This highland area is rich in mineral resources which tend to be localized. The newer Appalachians, as a whole, have the greatest bituminous coal reserve in the world and are the site of the first developed petroleum field in the United States. Elsewhere, minor coal deposits are found in Nova Scotia; iron ores in Newfoundland, New England, New Jersey, and Alabama; lead, zinc, wolframite in New Jersey; asbestos in Quebec; limestone, granite, marble, and slate in various localities; and gold in North Carolina. The Appalachian coal resources have contributed greatly to making the United States the world's leading industrial nation and locating much of this industry in the northeastern portion of the country.

New England–Maritime province. As a result in part of continental and mountain glaciation, but also because of extensive recent stream erosion, much of this province is rough land and some is more rugged than most of the Appalachians. Settlement has largely been limited to the coastal borders or the larger lowlands trapped inland among the highlands. Much of New England and Newfoundland is composed of crystalline rocks that have been subjected to considerable warpage, folding, uplift, and erosion, whereas the western part of New England and the Canadian section south of the St. Lawrence is made of sediments that were folded, compressed, and only locally metamorphosed into crystallines Brief notes on the eight commonly recognized subdivisions follow.

Newfoundland Ancient and Paleozoic crystalline rock structures of northeast trend compose this detached segment of the Appalachians The island is a tilted low plateau sloping generally to the east. A few peaks along the western edge have elevations of 2500 ft. The irregular coastline with long, narrow, parallel peninsulas results from the basic structural alignment and the intermixture of rocks with varying resistance to erosion Spruce forests of the interior support wood-pulp and paper-making industries, and high-grade iron ore is available on Bell Island in Conception Bay.

Coastal Border and Nova Scotia The Coastal Border and Nova Scotia section resembles Newfoundland in structure except that the orientation



; 5 Diagram of land surface and structural relationships in southern part of New England (From A. K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

more east-west and the rocks are geologically younger Very little of the area rises above 500 ft elevation so that local relief is moderate. Lowlands between the ridges in Nova Scotia provide better-quality soils which are the sites for agricultural settlements Coal, mined in Cape Breton Island is the base for a small-scale steel industry Underlying crystalline rocks, exposed along the ocean shoreline by wave cutting, provide a very irregular shore with small bays, headlands, and islands The coastal strip, which becomes narrower to the south, is covered by unconsolidated marine clays which are intensively farmed in New Brunswick and southern Maine but elsewhere have reverted to second-growth scrub forest after lumbering Fishing and boat building support numerous small settlements along the coast A few large cities such as Boston and Portland depend on trade and manufacture of raw materials imported from outside this section Glaciation contributed to the irregularity of the coastline The continental shelf here is the wide Grand Banks noted for fishing and for stormy seas.

New England Upland section. Old rocks of this area constitute a core of the New England Maritime province. The crystalline rocks have been subject to a number of peneplanations (*see* GEOMORPHOLOGY, FLUVIAL EROSION CYCLE) during and since Paleozoic time so that the area is a plateau with the hill crests varying from 800 to 2000 ft in elevation Beginning on the shore of Baie de Chaleur, the Upland extends south to Long Island Sound with a prong forming Manhattan Island and another protruding southwestward across the Hudson River to the vicinity of Reading, Pennsylvania The southern portions of this section are lower than the rest, ranging from sea level to elevations of a few hundred feet, and creating hilly belts rather than plateaus Forests are maintained in the central and northern sections while efforts to reforest the area are progressing in the south

Connec...

...the southern upland into two segments.

Taconic section This is a narrow strip lying between the New England Upland and the Green Moun...
Low...
sedi...
...

2000 ft elevation compared to the lower relief in the western Upland.

Green Mountain section. Some of Vermont and southern Quebec resembles the Taconic section in that it is a folded area with the axis running north-south, but the Green Mountain folding was more intensive so that most of the rocks are metamorphosed and intrusions of volcanics provide a greater variety of rocks. More resistant ridges reach 3000 ft elevation. Relatively little lowland is available for agriculture; the main activity is mining granite, marble, and, in Canada, asbestos.

White Mountain section. Here the surface rises above the New England Upland in northern New Hampshire and western Maine to an elevation of 6000 ft It trends in a northeasterly direction from the Green Mountains, but is composed of individual erosional remnants rather than parallel ridges. Higher elevations are above timber line and the heavy snows in this section have encouraged its use for winter sports.

Gaspé section This coastal upland resembles the Taconic Mountains except that the folds extend eastward around the north portion of the New England area to the Gulf of St. Lawrence Here elevations reach 4250 ft. The sparse population is concentrated in small villages along the coast.

Appalachian provinces. Northwest and west of the New England Maritime province are structures of folded Paleozoic rock in the St. Lawrence Valley ...

...their physical and related geographical characteristics are such that it is logical to consider them related parts of the northeast-southwest extent of the Newer Appalachians, variously called the folded valley and ridge province, the Great Valley, and numerous other names in local parts (such as Shenandoah Valley) The regionally parallel ridges and valleys are subdued, but present, in these northeastern parts of the Newer Appalachians, but the ridges of relatively resistant rock upstand in linear pattern of hill and low mountain terrain between Pennsylvania and Alabama

A sketch of geological history may be ...

...These beds extend great dis-

island waters of the Arctic archipelago. These areas are little-known. The southern part appears to be a poorly drained, low plateau with a thin cover of coniferous forest. In the north, crystalline rock slopes gently northward and forest gives way to tundra. These Barren Lands are the home of caribou and musk oxen. Some uranium is mined in the south, in Manitoba, and on the shores of Great Bear Lake in the north. Adequate prospecting should uncover additional mineral resources.

APPALACHIAN HIGHLANDS

A great variety of highlands and associated lowlands in eastern United States and southeastern Canada are considered as composing the Appalachian Highlands, making it a very complex region. It is generally divided into the New England-Maritime Canada province and the Appalachian province, which in turn is further subdivided into older and newer sections.

Rocks ranging from ancient crystalline to Recent sediments are found in the same geographic area in various sections of this highland region. The crystalline rocks predominate in the northern and eastern portions while the northwest and west are characterized by sediments of Paleozoic time. *See* GEOLOGICAL TIME SCALE

Geologic structures of almost every type, which have been subjected to a variety of erosional and depositional processes, are found somewhere in this region. Doming, folding, faulting, and tilting, as well as volcanic activities, contribute to the variety of surface features. Major uplifts followed by long periods of erosion have left what might otherwise have been a very high and rugged mountain area as a generally subdued highland with maximum elevation only slightly exceeding 6000 ft in a few summits. Even the peaks are smoothed and rounded by erosional forces. The northeast-southwest trend of the province from Newfoundland to northern Alabama appears in most of the individual structures.

The severe winter phase of the continental humid climate prevails throughout the area providing marked seasonal changes between winter and summer. The more elevated sections in the south have cooler temperatures than the latitudinal position would indicate, thus providing relief from summer heat in adjacent lowlands. Also, the uplands receive from 35 to 55 in. of precipitation which makes this a well-watered region and in winter provides a deep snow blanket.

Podzol and gray podzolitic soils prevail where slopes permit the accumulation of regolith (fragmental mineral material) and erosion has been less rapid than the soil-making processes. Where formed from crystallines, these soils are acidic in reaction but are neutral or alkaline in the limestone outcrops which provide the better soils for agriculture.

The mixed mid-latitude broad-leaf coniferous forest occurs throughout the region, but locally either the broad-leaf or coniferous species may occur in pure stands with different species predominating (Fig. 4). Excessive lumbering has changed the character of the remaining forest, even on the more rugged and less settled parts of the region.

The larger animals native to the Appalachians include bear and deer, and a few moose in the north, and considerable numbers of both larger and smaller animals still live in the area. Small game, fur bearers, and meat animals include beaver, fox, opossum, and squirrels.

The rivers are mostly small and turbulent. The relatively heavy rainfall supplies abundant water that drains out of the region and the upland character of the area provides many small hydroelectric-power sites. The water resources, however, have been useful in attracting industries that utilize large amounts of water, such as papermaking and the synthetic fiber industry.

This highland area is rich in mineral resources which tend to be localized. The newer Appalachians, as a whole, have the greatest bituminous coal reserve in the world and are the site of the first developed petroleum field in the United States. Elsewhere, minor coal deposits are found in Nova Scotia; iron ores in Newfoundland, New England, New Jersey, and Alabama; lead, zinc, wolframite in New Jersey; asbestos in Quebec; limestone, granite, marble, and slate in various localities; and gold in North Carolina. The Appalachian coal resources have contributed greatly to making the United States the world's leading industrial nation and locating much of this industry in the northeastern portion of the country.

New England–Maritime province. As a result in part of continental and mountain glaciation, but also because of extensive recent stream erosion, much of this province is rough land and some is more rugged than most of the Appalachians. Settlement has largely been limited to the coastal borders or the larger lowlands trapped inland among the highlands. Much of New England and Newfoundland is composed of crystalline rocks that have been subjected to considerable warpage, folding, uplift, and erosion, whereas the western part of New England and the Canadian section south of the St. Lawrence is made of sediments that were folded, compressed, and only locally metamorphosed into crystallines. Brief notes on the eight commonly recognized subdivisions follow.

Newfoundland. Ancient and Paleozoic crystalline rock structures of northeast trend compose this detached segment of the Appalachians. The island is a tilted low plateau sloping generally to the east. A few peaks along the western edge have elevations of 2500 ft. The irregular coastline with long, narrow, parallel peninsulas results from the basic structural alignment and the intermixture of rocks with varying resistance to erosion. Spruce forests of the interior support wood-pulp and paper-making industries, and high-grade iron ore is available on Bell Island in Conception Bay.

Coastal Border and Nova Scotia. The Coastal Border and Nova Scotia section resembles Newfoundland in structure except tha the orient

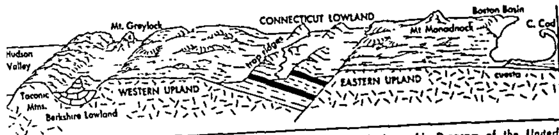


Fig 5 Diagram of land surface and structural relationships in southern part of New England (From A K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

is more east-west and the rocks are geologically younger Very little of the area rises above 500 ft in elevation so that local relief is moderate Lowlands between the ridges in Nova Scotia provide better-quality soils which are the sites for agricultural settlements Coal, mined in Cape Breton Island, is the base for a small-scale steel industry Underlying crystalline rocks, exposed along the ocean shoreline by wave cutting, provide a very irregular shore with small bays, headlands, and islands The coastal strip, which becomes narrower to the south, is covered by unconsolidated marine clays which are intensively farmed in New Brunswick and southern Maine but elsewhere have reverted to second-growth scrub forest after lumbering Fishing and boat building support numerous small settlements along the coast. A few large cities such as Boston and Portland depend on trade ... of raw materials imported from ... the ... shelf ... and for stormy seas

... Old rocks of this ...

PHOLOGY; FLUVIAL EROSION CYCLE) during and since Paleozoic time so that the area is a plateau with the hill crests varying from 800 to 2000 ft in elevation Beginning on the shore of Baie de Chaleur, the Upland extends south to Long Island Sound with a prong forming Manhattan Island and another protruding southwestward across the Hudson River to the vicinity of Reading, Pennsylvania The southern portions of this section are lower than the rest, ranging from sea level to elevations of a few hundred feet, and creating hilly belts rather than plateaus Forests are maintained in the central and northern sections while efforts to reforest the area are progressing in the south

Connecticut Lowland A down-faulted and tilted block of Triassic rocks through which the Connecticut River flows makes an area of fertile soil and divides the southern upland into two segments.

Taconic section This is a narrow strip lying between the New England Upland and the Green Mountains to the east and the Hudson-Champlain Lowland to the west (Fig 5). It is an area of folded sediments that have in part been metamorphosed creating slate from shales The higher ridges reach 2000 ft elevation compared to the lower relief in the western Upland

Green Mountain section. Some of Vermont and southern Quebec resembles the Taconic section in that it is a folded area with the axis running north-south, but the Green Mountain folding was more intensive so that most of the rocks are metamorphosed and intrusions of volcanics provide a greater variety of rocks. More resistant ridges ... 2000 ft elevation. Relatively little lowland is ... is min... ...tos.

White Mountain section Here the surface rises above the New England Upland in northern New Hampshire and western Maine to an elevation of 6000 ft It trends in a northeasterly direction from the Green Mountains, but is composed of individual erosional remnants rather than parallel ridges Higher elevations are above timber line and the heavy snows in this section have encouraged its use for winter sports.

Gaspé section. This coastal upland resembles the Taconic Mountains except that the folds extend eastward around the north portion of the New England area to the Gulf of St. Lawrence. Here elevations reach 4250 ft. The sparse population is concentrated in small villages along the coast.

Appalachian provinces. Northwest and west of the New England-Maritime province are structures of folded Paleozoic rock in the St. Lawrence Valley, the Champlain Lowland, and the Hudson Valley. Although their surface character and soils have been modified by recent (Pleistocene) glaciation, their physical and related geographical characteristics are such that it is logical to consider them related parts of the northeast-southwest extent of the Newer Appalachians, variously called the folded valley and ridge province, the Great Valley, and numerous other names in local parts (such as Shenandoah Valley). The regionally parallel ridges and valleys are subdued, but present, in these northeastern parts of the Newer Appalachians, but the ridges of relatively resistant rock upstand in a linear pattern of hill and low mountain terrain between Pennsylvania and Alabama

A sketch of geological history ...

great Paleozoic Appalachian highland and deposited into beds to the westward in what is now interior North America These beds extend great d

tances into the present interior plateaus and plains, but the deposition was particularly heavy in a long northeast-southwest geosyncline, or developing trough, just westward of the old eroding highland. Lateral pressure subsequently folded the sedimentary rock beds of the geosyncline into generally open folds. It is thought that these fold structures were eroded to a peneplain during the Mesozoic Era, and that later uplift or rejuvenated stream action is now etching the region at a differential rate —more rapidly in the less-resistant rocks. This leaves the edges of the more-resistant layers upstanding in the generally linear pattern of the Newer Appalachians. The summits of the more-resistant ridge structures may represent in part the past (Cretaceous) peneplain level, thereby explaining the rather common accordance of summit level of the ridges in many parts of the Newer Appalachians.

To the west today are the several sections of the Appalachian plateau and to the east the long-eroded remnants of the old Appalachian highlands—the Older Appalachian provinces. These extend alongside the Newer Appalachians, to form, with the New England-Maritime provinces, the present Appalachian Highlands of southeastern North America.

Older Appalachians provinces. Mostly crystalline rocks, igneous and metamorphic of long and complex history, underlie this two-part region of northeast-southwest trend between the New Appalachians province and the sedimentary rocks of the coastal plain. In a sense these are the stump and root rocks of the great Paleozoic Appalachian highland that presently appear at the face of the earth. The materials, long eroded from the overlying mass, were deposited not only in the beds of interior regions, such as the present Newer Appalachians and the Appalachian plateaus, but also especially in great beds that now comprise the structures of the coastal plain and the continental shelf. Thus the older crystalline rocks also form the base upon which the coastal-zone sediments rest Because the old-rock areas were not affected by Pleistocene glaciation, present surface of these provinces is generally deeply mantled by soil materials. In considerable contrast to the New England-Maritime provinces, rocky and bouldery soils are rare; bare rock exposures are uncommon enough even in mountainous parts to be conspicuously notable; and the interesting lakes of the New England area are lacking.

The two parts are (1) a long belt of the Blue Ridge section along the northwestern side, of hill and low mountain country overlooking the Newer Appalachians; and (2) the somewhat broader and longer Piedmont upland extending as a rolling and occasionally hilly country, mostly between the mountains of the Blue Ridge section and a lower eastern border with the coastal plain lowlands. Both of these sections extend from rather blunt ends in North Georgia and Alabama, respectively, to narrow prongs in Pennsylvania and New Jersey Between these prongs and two similar but southward-pointing prongs of the New England uplands stretches a Triassic lowland. This separates the New England provinces, at about the location of New York City, from the Older Appalachians in Virginia and to the southwest.

1. The F[illegible] mass of the Older [illegible] lowland fl[illegible] side and distinctly above the Piedmont upland on the Atlantic side. The Roanoke River, near Roanoke, Virginia, flows through a water gap from the valleys of the Newer Appalachians and thence eastward to the Atlantic, and the Blue Ridge is segmented to the north by the James and the Potomac Rivers passing eastward through similar water gaps. In all these parts the mountains vary from a fairly compact range of irregularly arranged rounded mountain summits (largely forested or reforested) to a definite linear ridge character in the northern parts. Today a growing proportion of this section is being preserved in national forests and parks Southward of Front Royal, Virginia, the Shenandoah National Park contains the scenic Skyline Drive for more than 100 mi to Rockfish Gap, and all parts are being used more and more for recreational activities, especially camping and hiking on the Appalachian Trail. Between Rockfish Gap and Roanoke the mountain highway, the Blue Ridge Parkway, runs through large areas of national forest. After a gap across the Roanoke Valley this great parkway resumes at the summit of the Blue Ridge

Southwest of the Roanoke River gap, the mountain country is perhaps better termed the Southern

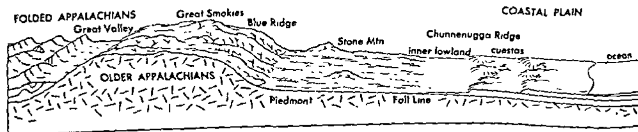


Fig 6. Structure and topography diagram of the Older Appalachians (Southern Appalachian Mountains, Piedmont plus bordering Coastal Plain) in southeastern United States. *(From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)*

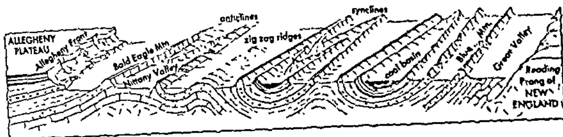


Fig 7. Surface and structure diagram of the Newer Appalachians (From A K Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

Appalachian Mountains For some 500 mi the Blue Ridge is the asymmetrical summit and stream divide of a scarp, narrow and steep in places, but up to 10 mi wide where it has been deeply scored by headwaters of streams flowing to the Atlantic. Occasional groups of hills and low mountains stand above the adjoining Piedmont as outlier remnants of the retreating scarp face. Beyond the elevated Blue Ridge Divide, headwaters of streams flowing to the Gulf appear to be eating out rolling to hilly floored upland basins interspersed between various masses and ranges of mountains reaching 2000 ft and more above the basins. A few peaks surmount the Blue Ridge summit, but more and higher rounded mountain summits rise beyond, Mt Mitchell at 6684 ft is the highest elevation east of the Mississippi River. All of the northwestern margins are mountainous in the Unaka Range and the Great Smoky Mountains, now a national park Here the strongly metamorphosed rocks of the Newer Appalachians are tightly compressed and occasionally pushed (overthrust) out over the normal sedimentary rocks of the adjoining valleys. The streams, cutting the basin floors, notch [illegible]

[illegible] some of the rainiest climate in the southeastern United States

A large part of the more rugged and forested terrain of this large mountainous highland is in various [illegible]

[illegible] of Asheville. Recreational uses increase with increased population and travel.

Although the mountains are much less of a barrier to travel and commerce now than in the past, here and in the Blue Ridge farther north roads, rail routes, and airlines still tend to follow certain routes of relative importance and easier passage. Travel other than local and recreational generally makes use of the adjoining valley or upland Piedmont routes for north south movement

2 The [illegible]
in the [illegible]

makes the Piedmont a major part of five southeastern seaboard states. The general regional slope of this old rock area is from northwest to southeast, and it is being drained and etched by streams flowing seaward from the Blue Ridge. Most parts of the area appear to have been reduced to a stage of old age (peneplain), and then, perhaps as a result of general upwarping, stream erosion becomes active again The present cycle of erosion is approaching its middle (submature) stage, so that the area is a rolling plain, with local relief in the steep-sided valleys seldom surpassing 100 ft. The inner Piedmont, toward the bordering mountains, is conspicuously marked by rougher hills and low mountain groups of monadnock remnants of once-higher overlying materials

Although some fine agriculture remains, the character of the soils and land use has altered from that of the past. Virtually all of the mixed forest has been cleared once and probably several times for cultivation of such crops as corn, tobacco, and cotton, as well as for general farming. With compact clay soils, heavy precipitation, and mild winters, erosion takes a heavy toll from the upper and generally more fertile parts of the soil, so that only with expensive methods of refertilization, erosion control, and special cropping techniques is agriculture maintained in competition with the interior plains or the newer lands of the Gulf Coastal Plain.

Much land is in regrowing timber, urban and industrial patterns are increasing Most volunteer regrowth is hardwood, but the greatest demand is for softwood timber and lumber. Some of the earlier cities have grown; most, however, are still small but increasing in number. Some of the cities established at the head of navigation at the fall line (or fall zone junction with the Coastal Plain structures) continue to grow (such as Philadelphia, Washington, Richmond, and Columbus, Ga.) but often for reasons other than tidewater navigation. Atlanta, at the southern end of the Blue Ridge section, is the largest industrial and commercial center.

Newer or Folded Appalachians. Five distinctive sections characterize the great northeast-southwest extent of these structures of eroded sedimentary folds: (1) St. Lawrence Valley section, (2) Champlain Lowland, (3) Hudson Valley section,

tances into the present interior plateaus and plains, but the deposition was particularly heavy in a long northeast-southwest geosyncline, or developing trough, just westward of the old eroding highland. Lateral pressure subsequently folded the sedimentary rock beds of the geosyncline into generally open folds. It is thought that these fold structures were eroded to a peneplain during the Mesozoic Era, and that later uplift or rejuvenated stream action is now etching the region at a differential rate —more rapidly in the less-resistant rocks. This leaves the edges of the more-resistant layers upstanding in the generally linear pattern of the Newer Appalachians. The summits of the more-resistant ridge structures may represent in part the past (Cretaceous) peneplain level, thereby explaining the rather common accordance of summit level of the ridges in many parts of the Newer Appalachians.

To the west today are the several sections of the Appalachian plateau and to the east the long-eroded remnants of the old Appalachian highlands—the Older Appalachian provinces. These extend alongside the Newer Appalachians, to form, with the New England-Maritime provinces, the present Appalachian Highlands of southeastern North America.

Older Appalachians provinces. Mostly crystalline rocks, igneous and metamorphic of long and complex history, underlie this two-part region of northeast-southwest trend between the New Appalachians province and the sedimentary rocks of the coastal plain. In a sense these are the stump and root rocks of the great Paleozoic Appalachian highland that presently appear at the face of the earth. The materials, long eroded from the overlying mass, were deposited not only in the beds of interior regions, such as the present Newer Appalachians and the Appalachian plateaus, but also especially in great beds that now comprise the structures of the coastal plain and the continental shelf Thus the older crystalline rocks also form the base upon which the coastal-zone sediments rest. Because the old-rock areas were not affected by Pleistocene glaciation, present surface of these provinces is generally deeply mantled by soil materials. In considerable contrast to the New England-Maritime provinces, rocky and bouldery soils are rare; bare rock exposures are uncommon enough even in mountainous parts to be conspicuously notable, and the interesting lakes of the New England area are lacking.

The two parts are (1) a long belt of the Blue Ridge section along the northwestern side, of hill and low mountain country overlooking the Newer Appalachians; and (2) the somewhat broader and longer Piedmont upland extending as a rolling and occasionally hilly country, mostly between the mountains of the Blue Ridge section and a lower eastern border with the coastal plain lowlands. Both of these sections extend from rather blunt ends in North Georgia and Alabama, respectively, to narrow prongs in Pennsylvania and New Jersey Between these prongs and two similar but southward-pointing prongs of the New England uplands stretches a Triassic lowland. This separates the New England provinces, at about the location of New York City, from the Older Appalachians in Virginia and to the southwest.

1. The Blue Ridge section is generally a mass of the Older Appalachian rocks upstanding above the lowland floors of the Newer Appalachians on one side and distinctly above the Piedmont upland on the Atlantic side. The Roanoke River, near Roanoke, Virginia, flows through a water gap from the valleys of the Newer Appalachians and thence eastward to the Atlantic, and the Blue Ridge is segmented to the north by the James and the Potomac Rivers passing eastward through similar water gaps. In all these parts the mountains vary from a fairly compact range of irregularly arranged rounded mountain summits (largely forested or reforested) to a definite linear ridge character in the northern parts Today a growing proportion of this section is being preserved in national forests and parks Southward of Front Royal, Virginia, the Shenandoah National Park contains the scenic Skyline Drive for more than 100 mi to Rockfish Gap, and all parts are being used more and more for recreational activities, especially camping and hiking on the Appalachian Trail. Between Rockfish Gap and Roanoke the mountain highway, the Blue Ridge Parkway, runs through large areas of national forest After a gap across the Roanoke Valley this great parkway resumes at the summit of the Blue Ridge

Southwest of the Roanoke River gap, the mountain country is perhaps better termed the Southern

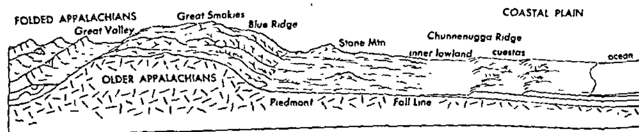
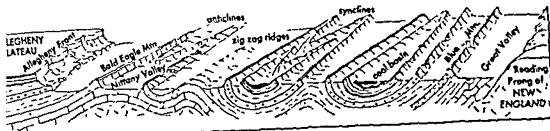


Fig. 6. Structure and topography diagram of the Older Appalachians (Southern Appalachian Mountains, Piedmont plus bordering Coastal Plain) in southeastern United States (From A. K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)



ɟ 7. Surface and structure diagram of the Newer ppalachians (From A K. Lobeck, *Physiographic Diagram of the United States, Hammond, 1957*)

ppalachian Mountains For some 500 mi the Blue Ridge is the asymmetrical summit and stream divide of a scarp, narrow and steep in places, but up o 10 mi wide where it has been deeply scored by headwaters of streams flowing to the Atlantic. Occasional groups of hills and low mountains stand above the adjoining Piedmont as outlier remnants of the retreating scarp face Beyond the elevated Blue Ridge Divide, headwaters of streams flowing to the Gulf appear to be eating out rolling to hilly floored upland basins interspersed between various masses and ranges of mountains reaching 2000 ft and more above the basins A few peaks surmount the Blue Ridge summit, but more and higher rounded mountain summits rise beyond, Mt Mitchell at 6684 ft is the highest elevation east of the Mississippi River All of the northwestern margins are mountainous in the Unaka Range and the Great Smoky Mountains, now a national park Here the strongly metamorphosed rocks of the Newer Appalachians are tightly compressed and occasionally pushed (overthrust) out over the normal sedimentary rocks of the adjoining valleys The streams,

fall rapidly out of the region The southern margins of the Southern Appalachian Mountains stand abruptly rugged above the Georgia Piedmont under some of the rainiest climate in the southeastern United States

A large part of the more rugged and forested terrain of this large mountainous highland is in various national forests and parks, but the rolling floors and basins are used for agriculture, habitation, industries dams and reservoirs, and the larger central city of Asheville Recreational uses increase with increased population and travel

Although the mountains are much less of a barrier to travel and commerce now than in the past, here and in the Blue Ridge farther north roads, rail routes, and airlines still tend to follow certain routes of relative importance and easier passage Travel other than local and recreational generally makes use of the adjoining valley or upland Pied-

makes the Piedmont a major part of five southeastern seaboard states. The general regional slope of this old-rock area is from northwest to southeast, and it is being drained and etched by streams flowing seaward from the Blue Ridge. Most parts of the area appear to have been reduced to a stage of old age (peneplain), and then, perhaps as a result of general upwarping, stream erosion becomes active again. The present cycle of erosion is approaching its middle (submature) stage, so that the area is a rolling plain, with local relief in the steep-sided valleys seldom surpassing 100 ft. The inner Piedmont, toward the bordering mountains, is conspicuously marked by rougher hills and low mountain groups of monadnock remnants of once-higher overlying materials.

Although some fine agriculture remains, the character of the soils and land use has altered from that of the past. Virtually all of the mixed forest has been cleared once and probably several times for cultivation of such crops as corn, tobacco, and cotton, as well as for general farming With compact clay soils, heavy precipitation, and mild winters, erosion takes a heavy toll from the upper and generally more fertile parts of the soil, so that only with expensive methods of refertilization, erosion control, and special cropping techniques is agriculture maintained in competition with the interior plains or the newer lands of the Gulf Coastal Plain.

Much land is in regrowing timber; urban and industrial patterns are increasing. Most volunteer regrowth is hardwood, but the greatest demand is for softwood timber and lumber. Some of the earlier cities have grown; most, however, are still small but increasing in number Some of the cities established at the head of navigation at the fall line (or fall zone junction with the Coastal Plain structures) continue to grow (such as Philadelphia, Washington, Richmond, and Columbus, Ga.) but often for reasons other than tidewater navigation. Atlanta, at the southern end of the Blue Ridge section, is the largest industrial and commercial center.

Newer or Folded Appalachians. Five distinctive sections characterize the great northeast-southwest extent of these structures of eroded sedimentary folds: (1) St. Lawrence Valley section, (2) Champlain Lowland, (3) Hudson Valley section

tances into the present interior plateaus and plains, but the deposition was particularly heavy in a long northeast-southwest geosyncline, or developing trough, just westward of the old eroding highland. Lateral pressure subsequently folded the sedimentary rock beds of the geosyncline into generally open folds. It is thought that these fold structures were eroded to a peneplain during the Mesozoic Era, and that later uplift or rejuvenated stream action is now etching the region at a differential rate —more rapidly in the less-resistant rocks. This leaves the edges of the more-resistant layers upstanding in the generally linear pattern of the Newer Appalachians. The summits of the more-resistant ridge structures may represent in part the past (Cretaceous) peneplain level, thereby explaining the rather common accordance of summit level of the ridges in many parts of the Newer Appalachians.

To the west today are the several sections of the Appalachian plateau and to the east the long-eroded remnants of the old Appalachian highlands—the Older Appalachian provinces. These extend alongside the Newer Appalachians, to form, with the New England-Maritime provinces, the present Appalachian Highlands of southeastern North America

Older Appalachians provinces. Mostly crystalline rocks, igneous and metamorphic of long and complex history, underlie this two-part region of northeast-southwest trend between the New Appalachians province and the sedimentary rocks of the coastal plain. In a sense these are the stump and root rocks of the great Paleozoic Appalachian highland that presently appear at the face of the earth. The materials, long eroded from the overlying mass, were deposited not only in the beds of interior regions, such as the present Newer Appalachians and the Appalachian plateaus, but also especially in great beds that now comprise the structures of the coastal plain and the continental shelf Thus the older crystalline rocks also form the base upon which the coastal-zone sediments rest Because the old-rock areas were not affected by Pleistocene glaciation, present surface of these provinces is generally deeply mantled by soil materials In considerable contrast to the New England-Maritime provinces, rocky and bouldery soils are rare; bare rock exposures are uncommon enough even in mountainous parts to be conspicuously notable, and the interesting lakes of the New England area are lacking.

The two parts are (1) a long belt of the Blue Ridge section along the northwestern side, of hill and low mountain country overlooking the Newer Appalachians; and (2) the somewhat broader and longer Piedmont upland extending as a rolling and occasionally hilly country, mostly between the mountains of the Blue Ridge section and a lower eastern border with the coastal plain lowlands. Both of these sections extend from rather blunt ends in North Georgia and Alabama, respectively, to narrow prongs in Pennsylvania and New Jersey Between these prongs and two similar but southward-pointing prongs of the New England uplands stretches a Triassic lowland. This separates the New England provinces, at about the location of New York City, from the Older Appalachians in Virginia and to the southwest.

1. The Blue Ridge section is generally a mass of the Older Appalachian rocks upstanding above the lowland floors of the Newer Appalachians on one side and distinctly above the Piedmont upland on the Atlantic side. The Roanoke River, near Roanoke, Virginia, flows through a water gap from the valleys of the Newer Appalachians and thence eastward to the Atlantic, and the Blue Ridge is segmented to the north by the James and the Potomac Rivers passing eastward through similar water gaps. In all these parts the mountains vary from a fairly compact range of irregularly arranged rounded mountain summits (largely forested or reforested) to a definite linear ridge character in the northern parts Today a growing proportion of this section is being preserved in national forests and parks. Southward of Front Royal, Virginia, the Shenandoah National Park contains the scenic Skyline Drive for more than 100 mi to Rockfish Gap, and all parts are being used more and more for recreational activities, especially camping and hiking on the Appalachian Trail. Between Rockfish Gap and Roanoke the mountain highway, the Blue Ridge Parkway, runs through large areas of national forest After a gap across the Roanoke Valley this great parkway resumes at the summit of the Blue Ridge

Southwest of the Roanoke River gap, the mountain country is perhaps better termed the Southern

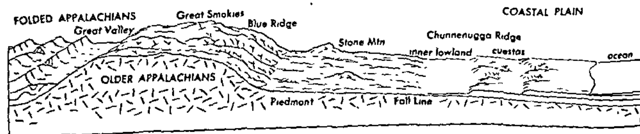


Fig. 6. Structure and topography diagram of the Older Appalachians (Southern Appalachian Mountains, Piedmont plus bordering Coastal Plain) in southeastern United States. (From A. K Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

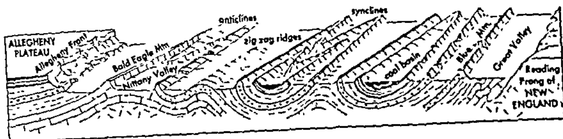


Fig 7. Surface and structure diagram of the Newer Appalachians. (From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

Appalachian Mountains. For some 500 mi the Blue Ridge is the asymmetrical summit and stream divide of a scarp, narrow and steep in places, but up 10 mi wide where it has been deeply scored by headwaters of streams flowing to the Atlantic. Occasional groups of hills and low mountains stand above the adjoining Piedmont as outlier remnants of the retreating scarp face. Beyond the elevated Blue Ridge Divide, headwaters of streams flowing to the Gulf appear to be eating out rolling to hilly floored upland basins interspersed between various masses and ranges of mountains reaching 2000 ft and more above the basins. A few peaks surmount the Blue Ridge summit, but more and higher rounded mountain summits rise beyond. Mt. Mitchell at 6684 ft is the highest elevation east of the Mississippi River. All of the northwestern margins are mountainous in the Unaka Range and the Great Smoky Mountains, now a national park. Here the strongly metamorphosed rocks of the Newer Appalachians are tightly compressed and occasionally pushed (overthrust) out over the normal sedimentary rocks of the adjoining valleys. The streams, cutting the basin floors, notch through the Unakas, but their upland gradients are upheld by passing over resistant rock sills, below which these streams fall rapidly out of the region. The southern margins of the Southern Appalachian Mountains stand abruptly rugged above the Georgia Piedmont under some of the rainiest climate in the southeastern United States.

A large part of the more rugged and forested terrain of this large mountainous highland is in various national forests and parks, but the rolling floors and basins are used for agriculture, habitation, industries, dams and reservoirs, and the larger central city of Asheville. Recreational uses increase with increased population and travel.

Although the mountains are much less of a barrier to travel and commerce now than in the past, here and in the Blue Ridge farther north roads, rail routes, and airlines still tend to follow certain routes of relative importance and easier passage. Travel other than local and recreational generally makes use of the adjoining valley or upland Piedmont routes for north-south movement.

2. The Piedmont section, much generalized about in the past, needs considerable reappraisal today. Its long extent, from New Jersey to Alabama, makes the Piedmont a major part of five southeastern seaboard states. The general regional slope of this old rock area is from northwest to southeast, and it is being drained and etched by streams flowing seaward from the Blue Ridge. Most parts of the area appear to have been reduced to a stage of old age (peneplain), and then, perhaps as a result of general upwarping, stream erosion becomes active again. The present cycle of erosion is approaching its middle (submature) stage, so that the area is a rolling plain, with local relief in the steep-sided valleys seldom surpassing 100 ft. The inner Piedmont, toward the bordering mountains, is conspicuously marked by rougher hills and low mountain groups of monadnock remnants of once-higher overlying materials.

Although some fine agriculture remains, the character of the soils and land use has altered from that of the past. Virtually all of the mixed forest has been cleared once and probably several times for cultivation of such crops as corn, tobacco, and cotton, as well as for general farming. With compact clay soils, heavy precipitation, and mild winters, erosion takes a heavy toll from the upper and generally more fertile parts of the soil, so that only with expensive methods of refertilization, erosion control, and special cropping techniques is agriculture maintained in competition with the interior plains or the newer lands of the Gulf Coastal Plain.

Much land is in regrowing timber; urban and industrial patterns are increasing. Most volunteer regrowth is hardwood, but the greatest demand is for softwood timber and lumber. Some of the earlier cities have grown; most, however, are still small but increasing in number. Some of the cities established at the head of navigation at the fall line (or fall zone junction with the Coastal Plain structures) continue to grow (such as Philadelphia, Washington, Richmond, and Columbus, Ga.) but often for reasons other than tidewater navigation. Atlanta, at the southern end of the Blue Ridge section, is the largest industrial and commercial center.

Newer or Folded Appalachians. Five distinctive sections characterize the great northeast-southwest extent of these structures of eroded sedimentary folds: (1) St. Lawrence Valley section, (2) Champlain Lowland, (3) Hudson Valley section,

tances into the present interior plateaus and plains, but the deposition was particularly heavy in a long northeast-southwest *geosyncline, or developing trough*, just westward of the old eroding highland. Lateral pressure subsequently folded the sedimentary rock beds of the geosyncline into generally open folds. It is thought that these fold structures were eroded to a peneplain during the Mesozoic Era, and that later uplift or rejuvenated stream action is now etching the region at a differential rate —more rapidly in the less-resistant rocks. This leaves the edges of the more-resistant layers upstanding in the generally linear pattern of the Newer Appalachians. The summits of the more-resistant ridge structures may represent in part *the past (Cretaceous) peneplain level*, thereby explaining the rather common accordance of summit level of the ridges in many parts of the Newer Appalachians.

To the west today are the several sections of the Appalachian plateau and to the east the long-eroded remnants of the old Appalachian highlands—the Older Appalachian provinces. These extend alongside the Newer Appalachians, to form, with the New England-Maritime provinces, the present Appalachian Highlands of southeastern North America.

Older Appalachians provinces. Mostly crystalline rocks, igneous and metamorphic of long and complex history, underlie this two-part region of northeast-southwest trend between the New Appalachians province and the sedimentary rocks of the coastal plain In a sense these are the stump and root rocks of the great Paleozoic Appalachian highland that presently appear at the face of the earth. The materials, long eroded from the overlying mass, were deposited not only in the beds of interior regions, such as the present Newer Appalachians and the Appalachian plateaus, but also especially in great beds that now comprise the structures of the coastal plain and the continental shelf Thus the older crystalline rocks also form the base upon which the coastal-zone sediments rest. Because the old-rock areas were not affected by Pleistocene glaciation, present surface of these provinces is generally deeply mantled by soil materials. In considerable contrast to the New England-Maritime provinces, rocky and bouldery soils are rare; bare rock exposures are uncommon enough even in mountainous parts to be conspicuously notable, and the interesting lakes of the New England area are lacking.

The two parts are (1) a long belt of the Blue Ridge section along the northwestern side, of hill and low mountain country overlooking the Newer Appalachians; and (2) the somewhat broader and longer Piedmont upland extending as a rolling and occasionally hilly country, mostly between the mountains of the Blue Ridge section and a lower eastern border with the coastal plain lowlands. Both of these sections extend from rather blunt ends in North Georgia and Alabama, respectively, to narrow prongs in Pennsylvania and New Jersey Between these prongs and two similar but southward-pointing prongs of the New England uplands stretches a Triassic lowland. This separates the New England provinces, at about the location of New York City, from the Older Appalachians in Virginia and to the southwest.

1. The Blue Ridge section is generally a mass of the Older Appalachian rocks upstanding above the lowland floors of the Newer Appalachians on one side and distinctly above the Piedmont upland on the Atlantic side. The Roanoke River, near Roanoke, Virginia, flows through a water gap from the valleys of the Newer Appalachians and thence eastward to the Atlantic, and the Blue Ridge is segmented to the north by the James and the Potomac Rivers passing eastward through similar water gaps. In all these parts the mountains vary from a fairly compact range of irregularly arranged rounded mountain summits (largely forested or reforested) to a definite linear ridge character in the northern parts Today a growing proportion of this section is being preserved in national forests and parks. Southward of Front Royal, Virginia, the Shenandoah National Park contains the scenic Skyline Drive for more than 100 mi to Rockfish Gap, and all parts are being used more and more for recreational activities, especially camping and hiking on the Appalachian Trail. Between Rockfish Gap and Roanoke the mountain highway, the Blue Ridge Parkway, runs through large areas of national forest. After a gap across the Roanoke Valley this great parkway resumes at the summit of the Blue Ridge.

Southwest of the Roanoke River gap, the mountain country is perhaps better termed the Southern

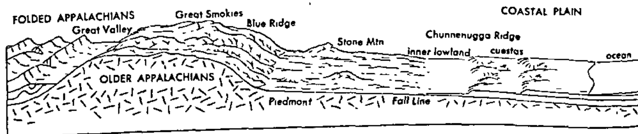


Fig. 6. Structure and topography diagram of the Older Appalachians (Southern Appalachian Mountains, Piedmont plus bordering Coastal Plain) in southeastern United States. (From A. K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

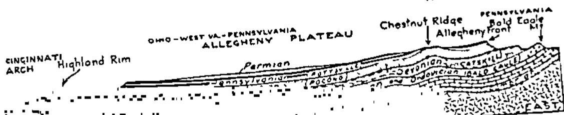


Fig 9 Structural section through the Appalachian Plateau (From A K. Lobeck, Geologic Map of the United States, Hammond, 1941)

The region is threaded with good roads and is progressively more densely settled toward New York City. The daily commuter movement adds tremendously to traffic, already heavy in this transit area in the hinterland of New York City. Some of the cities of this province are parts of the industrial and commercial complex, those on tidewater also have contact with ocean shipping Many of the cities, however, simply serve as suburban residential areas

Appalachian plateau Little-disturbed rock layers, Paleozoic sediments offscoured from the great Paleozoic highlands of Appalachia and thinning from east to west, upstand in a platterlike structure to form with local variations, the westernmost Appalachian provinces The great, slightly down-bent basin or synclinal structure contains some minor folds, especially along the eastern part and where this province adjoins, with an abrupt front the tighter folds of the Newer Appalachians province The rough easterly front is commonly designated Allegheny (in the north) and Cumberland "mountains" or front (in the south) Variably rough fronts also border the Mohawk Valley and Lake Erie lowland on the north and characterize the western margins from New York State into Alabama where these structures are overlain by those of the coastal plain on the south As a result of past peneplanation of the warped structure, older layers of rock underlie the surface near the margins and younger layers are preserved in more central portions of the plateau The great Appalachian coal fields in some of the younger rocks (Permian and Pennsylvanian) are thus largely in the central portions of the plateau especially in Pennsylvania West Virginia, and eastern Kentucky, whereas the local folds contained the petroleum deposits which once made the plateau an important petroleum region

Some structural differences and variations in erosion contribute to contrast in mature erosional dissection, the resulting landforms with associated features can be divided into five distinctive parts

1 Catskill section An abrupt stepped escarpment rises some 3000 ft from the western side of the Hudson Valley in southeastern New York as the front for an area of massive sandstone now maturely dissected into what are commonly called the Catskill Mountains Valleys are cut as much as 2000 ft beneath the general upland level, remaining at about 4000 ft in rounded to flattish summits. The coarse-textured bold terrain was somewhat marked by recent glaciation but contains few lakes. The forested land is used for recreation and water supply reservoirs for New York City, some 80 mi distant. The somewhat circular area stands, in all directions except the highest east, above the lower plateau lands of southern New York State and northeastern Pennsylvania.

2. New York State section Maturely dissected ... recent glacia-

west Allegheny escarpment stands above the Mohawk depression and the lowlands of Lakes Ontario and Erie In western New York State several of the valleys of this northern margin were occupied and enlarged by tongues of glacial ice, resulting in the Finger Lakes of the present landscape (see FINGER LAKES). Mostly cleared of forest and less intensively settled and developed than the Great ... lowlands on the north this un-

few through routes that are important for travel and communication

3 Allegheny section This unglaciated plateau area of western Pennsylvania and parts of Maryland and West Virginia contains such a deeply and intricately dissected hill and low mountain terrain that it is commonly termed Allegheny Mountains. The wild and rugged eastern Allegheny front is ascended by the Baltimore and Ohio and the Pennsylvania Railroads by spectacular horseshoe curves. Some of the locally anticlinal structures once contained oil fields. The rough land still contains large forest areas, scattered bituminous coal mining and associated industries, little agriculture, small population, and few towns or industries

4 Kanawha section. With local relief up to 1000 ft or more, the Kanawha section presents an intricately etched, rough hill and low mountain country in most of West Virginia and eastern Ohio. The Kanawha River flows in a gorge northwestward toward the Ohio River valley to form a constricted but important way through the plateau The Chesapeake and Ohio Railway and U.S Highway 60 with numerous mining communities and associated industries are narrowly strung along the valley. The highland remains more forested than cleared for agriculture although sheep grazing is developed to

(4) Pennsylvania-Virginia section, and (5) Tennessee Valley section. All of these have considerable significance in relation to past and current routes of travel—especially in the axial direction (NE-SW) of the upstanding ridges as well as the connecting lowlands developed on the weaker rocks of the folded structures. The first three form an easy route from eastern United States to Canada. The Hudson Valley connects at Albany with the Mohawk depression to make an easy route from New York City to the great interior plains via the Great Lakes region. This contributed much to the growth and dominance of New York City during the settlement and development of the country, and the route still has multiple "water-level" railway and highway lines. Predominant physical character of the first three was discussed at the beginning of the section on the Appalachian provinces. Major characteristics of the fourth and fifth sections are outlined below.

The classic development of the Appalachian type of peneplaned and rejuvenated folded mountain and valley terrain is found in the broad (75 mi) band across Pennsylvania (see Fig. 7) This continues, in a narrower band, in West Virginia and western Virginia. The peculiar pattern of upstanding linear ridges would be difficult to cross transversely, were it not for several rivers cutting through water gaps to cross the region. Examples are the Delaware and Susquehanna Rivers of Pennsylvania and the Potomac and New Rivers farther southward, these gaps create easier transverse rail and highway routes.

Some of the sedimentary rocks result in interesting features or contain valuable minerals. In being folded, heat and pressure altered to hard coal (or anthracite) some coal beds in parts of Pennsylvania which are soft (or bituminous coal) in the less-disturbed beds of the plateaus to the west (Fig. 7). Oil fields are developed in some parts of the area in Virginia and West Virginia. Some of the limestones, underlying the valley floors because they are a less-resistant rock under humid climate, develop typical karst features of sinkholes and many caverns, such as those found in western Virginia (see KARST TOPOGRAPHY). Much of the hilly and mountainous ridge lands is regrowing forest in national forests, but many of the valley portions are still agricultural areas The lowland portions contain important northeast-southwest routes of travel and commerce and scattered towns and small cities.

The Tennessee Valley section of the Newer or Folded Appalachians presents several aspects differing from the more northern parts. Folding of the rock layers is less open, and a series of thrusts from the east results in lower, repeated, and more numerous ridges and intervening valleys. The trellised pattern of valleys and streams tributary to the Tennessee River contains numerous sites of note in the regional program of the TVA. Available power and associated resources have contributed to the growth of cities such as Bristol, Knoxville, and Chattanooga. The most southern parts of this valley and ridge section, drained largely by the Coosa River system, constitute a district noted for iron and steel industries, centered at Birmingham. In close proximity to the fine coal of the adjoining plateau, the easily available iron ore deposits and fluxing limestone of the valley and ridge structures have been important in the development of the industry of this area.

Triassic Lowland. A small but distinctive and useful area makes a strategically located, primarily lowland province amid the Appalachian Highlands. Southwestward from the Hudson River, this Triassic Lowland extends from between the Manhattan prong and longer Reading prong of the New England Upland area in a curving band across northern New Jersey. Between suburban New York and Trenton, the lowland borders the inner Coastal Plain to give easy ways inland beyond the northward-protruding Trenton prong of the Appalachian Piedmont. These ways lead to the northeast-southwest valley routes or to the transverse gapways, as in the Susquehanna Valley near Harrisburg, in the Newer Appalachians. The Triassic Lowland, however, swings in a narrowing band southward in Pennsylvania across Maryland and into Virginia between the Trenton Piedmont prong and the Carlisle or Cumberland prong of the Blue Ridge southward of Harrisburg.

The westward-dipping shale and sandstone layers of the area are eroded to low hills and plains, but occasionally interbedded igneous rocks remain more upstanding at the surface in a few ridges, especially the Wachungs, and in the trap-rock cliffs of the Palisades along the Hudson River The gray igneous rocks are used as crushed-rock road material, and the softer red sandstones provided materials for buildings of Dutch settlers and for the "brownstone" fronts in older residential districts of New York City.

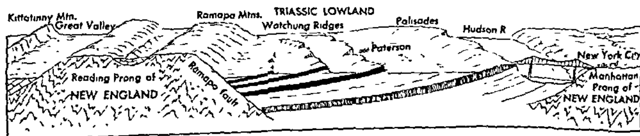


Fig. 8 Diagram of structure and surface patterns in ... K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

Proximity to great urban areas is reflected from Long Island through New Jersey and eastern Maryland Truck farming utilizes many areas of sandy soils, and the beaches include some of the most outstanding coastal resorts in the United States. Much of the more southern portions is returning to forest uses along with continuing truck farming and specialized agriculture Along with scattered urban population, some of the great urban areas center upon the older seaports, such as New York and Baltimore at the fall line

Sea Island section. The Coastal Plain in South Carolina and Georgia is broad and comparatively simple in pattern The fall zone is a nearly indistinguishable transition marked by low hills and several cities, such as Columbus, Macon, and Augusta, Georgia, and Columbia, South Carolina Regrowing forests, grazing, and agriculture including early-season truck farming utilize much of this area. Wet lands and tidewater intermingle in swamp forest, marsh, and muddy or sandy islands in the outer Coastal Plain. The name Sea Island is taken from the repeated coastal pattern of longshore islands The outer lowland plain contains a few urban foci at seaport locations such as Charleston [illegible]

[illegible]

depressions, underground drainage, and great springs, developed in the limestone rock Marine organic sediments yield valuable phosphate materials Much of the rural land of the north is taken up by piney forest, considerable grazing, and citrus fruit raising A great deal of the land of southern Florida is very low, youthfully exposed plains surface Lake Okeechobee and the poorly drained Everglades are examples of initially undrained depressions in such a land surface, but artificial drainage projects are reclaiming more land for fruit and truck farming for off season and other marketing in northern United States

Tropical mild winters and attractive coastal-zone patterns stimulate a variety of developments. Jacksonville and Miami are examples of growing seaports and industrial places as well as resort centers Older shell beach ridges, lagoons, and current barrier beaches comprise much of the eastern shore with its speculative and resort developments This coastal line, partly barrier bar and partly coral in origin swings southward in the chain of the Keys to the marine base at Key West. Offshore to the southeast a broad protrusion of continental shelf is marked by the Bahama Islands in a zone of recent prospecting for petroleum Western Florida coastal zones resemble the other Gulf sections

East and West Gulf sections. The Mississippi alluvial plain divides these two coastal plain segments bordering the Gulf of Mexico. The eastern part, in Alabama and Mississippi, presents the typical land features of the belted coastal plain with cuestaform ridges separating intervale and inner lowlands The previously rich agricultural land of the inner lowland in the Alabama Black Belt is faced with problems of depletion of soil fertility under present plans of use and from competition with cultivation farther west and in the interior plains. As well as continuing agriculture, these coastal plain lands support grazing in forest and grassy openings, and the piney woods are regrowing as sources of lumber and turpentine. Some petroleum fields are producing from local domes, folds, and entrapping structures within the coastal plain rocks. Small cities and towns are mostly scattered [illegible]

[illegible] bama.

The West Gulf Coastal Plain continues from the Mississippi lowlands to the Rio Grande as a wide belted coastal plain but with certain marked differences. The fertile inner lowland terminates against an abrupt fault scarp, most of which is the Balcones Escarpment [illegible] ern [illegible] by th [illegible] by th [illegible] San Antonio The outer Coastal Plain, with barrier beaches, lagoons, estuarial bays, and marshes, maintains channeled seaport contact through Galveston, Houston, Beaumont, Port Arthur, and Corpus Christi, which are also growing industrial places Notable spots of subtropical agriculture now mark the coastal zone in machine cultivation of rice, citrus and other fruits, and truck raising, especially citrus in the lower Rio Grande vicinity.

Petroleum fields have been prospected with geophysical and other techniques to develop some of the continent's outstanding gas and oil productions [illegible]

[illegible] petroleum exploitation into the submarine sediments offshore in the continental shelf. This happened first in California, then in the Gulf of Mexico, and is recently showing promise of future success in the Bahama section and possibly in the Atlantic continental shelf off the eastern United States

The Gulf Coastal Plain narrows southward in Mexico, is interrupted in several places, and becomes too hot and steaming a tropical lowland in Mexico to attract much settlement. Tampico and Veracruz are seaport cities with rail and highway connection into highland Mexico Petroleum production declined in this area, but it is now being revived on a considerable scale. The volcanic highlands of Mexico end in abrupt slopes to the sea northward of Veracruz, and the striking Tuxtlas volcanoes rise on the seaward margin of the coastal plain to the south of that city. Beyond, at the Isthmus of Tehuantepec, the coastal plains of the Gulf of Mexico and of the Pacific Gulf of Tehuantepec are nearly joined.

Yucatan section. More coral than limestone rock material, more definitely tropical climate and veg-

some extent. Petroleum extraction continues in several parts. The central portion is in the heart of the Appalachian bituminous coal fields and coal industries. The great urban center of Pittsburgh stands within the region at the head of the Ohio Valley. There coal, coke from neighboring industries, and various iron ores have been combined to make this a leading iron and steel manufacturing center. Other chief centers of population in the valleys and a few such as Charleston and Wheeling, West Virginia, have become sizable urban centers.

5. Cumberland section. In this southern portion the plateau belt narrows in width and gradually decreases in local relief. The breadth of the generally slight synclinal structure is broken by a few longitudinal folds. These anticlinal parts are etched out as valleys between large, long, flat-topped remnants of the plateau upland, such as Sand Mountain and West Sand Mountain. Lookout Mountain is a similar remnant, judged by some to belong to the Folded Appalachians and by others to be part of the plateau system. The Tennessee River cuts an irregular course across the region and forms many valley sections now developed by dams and reservoirs as important links in the TVA system. Farther north the eastern part of the plateau, called the Cumberland Mountains, contains historic Cumberland Gap. This is actually a narrow valley ascending the eastern margin of the upland near the border junction of Virginia, Kentucky, and Tennessee, and is a route still used by railroads and highways.

Fine coal fields have been developed in the Kentucky, Tennessee, and Alabama parts of the plateau region. The Warrior coal field contributes largely to the great iron and steel industries centering in Birmingham, Alabama.

SOUTHEASTERN COASTAL PLAIN

A distinctive coastal plain margins North America, along the coastal zone of the Atlantic Ocean and the Gulf of Mexico, in varying widths but nearly unbroken from Cape Cod and Long Island to Yucatan in southeastern Mexico. In most parts, very flat seaward margins of low-lying and most recent sedimentary deposits are followed inland by a tendency for a coastwise banding of features. A landward increase in elevation and erosional dissection also tends to contribute toward rolling and irregular plains with local relief up to a very few hundreds of feet. A gradual gentle uplift or upwarp of the geologically recent sedimentary rock layers means that they are little disturbed from their original horizontal attitude and dip slightly seaward and that such regions have most uplift inland toward bordering backland areas of differing character. Erosion, therefore bares lower and somewhat older sedimentary layers progressively in a landward direction. Landward-facing cuestas or bands of low hills may develop on more-resistant rock exposures; whereas lower relief and occa- [...] develop by differential erosion [...]taceous clays, are termed inner lowland (*see* ESCARPMENT). The landward coastal plain boundary is commonly the inner margin of the Cretaceous clays, and the border between the coastal plain layers and the oldland or backland rocks is known as the fall line or fall zone. *See* FALL LINE.

Several other factors influence differences within the great extent of these southeastern coastal plains. Recent changes in relative elevation of land and sea constitute such an influence. A predominance of land subsidence marks the region from North Carolina to Cape Cod with many embayments and shorelines irregular in contrast to most other parts, of recent upraisal or stillstand of the land relative to the sea. A strongly marked variance from north to south in climatic characteristics with an especially steep temperature gradient is reflected in such aspects as natural vegetation (Fig 4) and length of growing season and agricultural production, as well as the pattern contrasts in winter resorts of Florida and the Gulf coast versus the definite winter influences of the New York and New Jersey coastal zones. *See* AIR TEMPERATURE; SOIL; SOIL, ZONAL DISTRIBUTION; VEGETATION ZONES (WORLD). Differing widths of valley bottom alluvium appear along streams and are a large and significant factor in case of valleys of trunk streams following the regional slope from the backland to the sea—the Apalachicola, Rio Grande, and particularly the great valley and delta of the Mississippi. *See* COASTAL PLAIN; DELTA; FLOOD PLAINS; FLUVIAL EROSION CYCLE; FLUVIAL EROSION LANDFORMS.

Within the essential unity of this great coastal plain form and structure, six or seven subdivisions are useful in outlining the regional variations characterizing its long extent: Northern embayed, Sea Island, Peninsular Florida, East Gulf section, Mississippi alluvial plain, West Gulf, and Yucatan section.

Northern embayed section. Considerable evidence indicates that the coastal plain features northward from Cape Fear and the estuary of the Cape Fear River, although affected by both emergence and submergence, are most marked from a recent predominance of submergence in relation to tidewater. The outer coastal plain flats are not only marked by swamps and marshes but tidewater extends upvalley inland in the North Carolina sounds and in the drowned-valley estuaries of Chesapeake and Delaware Bays. Bars and barrier beaches form on the longshore margin of the sediments extending beneath the shallow sea, partially enclosing the lagoons and other tidewater features of this coastal plain section. *See* COASTAL LANDFORMS; MARINE MARSH, SHORE PROCESSES.

From Long Island to Cape Cod, subsidence makes most of the recent coastal plain, including the inner lowland (as in Long Island Sound), now a part of the continental shelf. This portion was glaciated, and moraines augment the cuestas and erosional remnants of the coastal plain that remain

Proximity to great urban areas is reflected from *Long Island through New Jersey and eastern Mary-*land. Truck farming utilizes many areas of sandy soils, and the beaches include some of the most outstanding coastal resorts in the United States Much of the more southern portions is returning to forest uses along with continuing truck farming and specialized agriculture Along with scattered urban population, some of the great urban areas center upon the older seaports, such as New York and Baltimore at the fall line

Sea Island section. The Coastal *Plain* in South Carolina and Georgia is broad and comparatively simple in pattern The fall zone is a nearly indis-

ing forests, grazing, and agriculture including early-season truck farming utilize much of this area Wet lands and tidewater intermingle in swamp forest, marsh, and muddy or sandy islands *in the outer* Coastal *Plain* The name Sea Island is taken from the repeated coastal pattern of longshore islands The outer lowland *plain contains a* few urban foci at seaport locations such as

depressions, underground drainage, and *great* springs, developed in the *limestone* rock Marine organic sediments yield valuable *phosphate* materials Much of the rural land of the north is taken up by piney forest, considerable grazing, and citrus fruit raising A great deal of the *land of southern Florida* is very low, youthfully exposed plains surface Lake Okeechobee and the poorly drained Everglades are examples of initially undrained *de*pressions in such a land surface, but artificial drainage projects are reclaiming more land for fruit and truck farming for off-season *and other* marketing in northern United States

Tropical mild winters and attractive coastal-zone patterns stimulate a variety of developments Jacksonville and Miami are examples of growing seaports and industrial places as well as resort centers. Older shell beach ridges, lagoons, and current barrier beaches comprise much of the eastern shore with its *speculative* and resort developments This coastal line, *partly barrier* bar and partly coral in origin, swings southward *in* the chain of the Keys to the marine base at Key West Offshore to the southeast a broad protrusion of continental shelf is marked by the *Bahama Islands* in a

two coastal plain segments bordering the Gulf of Mexico. The eastern part, in Alabama and Mississippi, *presents* the typical land features of the belted coastal plain with cuestaform ridges separating intervale and inner lowlands The previously rich agricultural land of the inner lowland in the Alabama Black Belt is faced with problems of depletion of soil fertility under present plans of use and from competition with cultivation farther west and in the interior plains. As well as continuing agriculture, these coastal plain lands support grazing in forest and grassy openings, and the piney woods are regrowing as sources of lumber and turpentine. Some petroleum fields are producing from local domes, folds, and entrapping structures within the coastal plain rocks. Small cities and towns are mostly scattered in the inland lowland belts except for a few in the outer coastal plain, such as Tallahassee and Pensacola in Florida and Mobile, Alabama.

The West Gulf Coastal Plain continues *from the* Mississippi lowlands to the Rio Grande as a wide belted coastal plain but with certain marked *differences*. The fertile inner lowland *terminates* against an abrupt fault scarp, *most of which is the* Balcones Escarpment up to the high plains of western Texas Eastward the Coastal Plain *is bordered by the Ouachita upland*. The inner belt is marked by the cities of Dallas, Fort Worth, Austin, and San Antonio The outer Coastal Plain, with *barrier* beaches, lagoons, *estuarial* bays, and marshes, maintains channeled

veston,

Corpus

places Notable spots of subtropical agriculture now mark the coastal zone in

of rice,

especiall

Petroleum fields have been prospected with geophysical and other techniques to develop some of *the continent's outstanding* gas and oil productions

first in California, *then in the Gulf* of Mexico, and is recently showing promise of future success in the Bahama section and *possibly* in the Atlantic continental shelf off the eastern United States

The Gulf *Coastal* Plain narrows southward in *Mexico*, is interrupted in several places, and becomes too hot and steaming a tropical lowland *in* Mexico to attract much settlement. Tampico and Veracruz are seaport cities with rail and highway connection *into highland Mexico* Petroleum production declined *in this* area, but it is now being *revived* on *a considerable* scale. The volcanic highlands of Mexico end in abrupt slopes to the sea northward *of* Veracruz, and the striking Tuxtlas volcanoes rise on the seaward margin of the coastal plain to the south of that city Beyond, at the Isthmus of Tehuantepec, the coastal plains of the Gulf of Mexico and of the Pacific Gulf of Tehuantepec are nearly joined.

Yucatan section. More *coral* than limestone rock material, more definitely tropical *climate* and veg-

some extent. Petroleum extraction continues in several parts. The central portion is in the heart of the Appalachian bituminous coal fields and coal industries. The great urban center of Pittsburgh stands within the region at the head of the Ohio Valley. There coal, coke from neighboring industries, and various iron ores have been combined to make this a leading iron and steel manufacturing center. Other chief centers of population in the valleys and a few such as Charleston and Wheeling, West Virginia, have become sizable urban centers.

5. Cumberland section. In this southern portion the plateau belt narrows in width and gradually decreases in local relief. The breadth of the generally slight synclinal structure is broken by a few longitudinal folds. These anticlinal parts are etched out as valleys between large, long, flat-topped remnants of the plateau upland, such as Sand Mountain and West Sand Mountain Lookout Mountain is a similar remnant, judged by some to belong to the Folded Appalachians and by others to be part of the plateau system. The Tennessee River cuts an irregular course across the region and forms many valley sections now developed by dams and reservoirs as important links in the TVA system. Farther north the eastern part of the plateau, called the Cumberland Mountains, contains historic Cumberland Gap. This is actually a narrow valley ascending the eastern margin of the upland near the border junction of Virginia, Kentucky, and Tennessee, and is a route still used by railroads and highways.

Fine coal fields have been developed in the Kentucky, Tennessee, and Alabama parts of the plateau region The Warrior coal field contributes largely to the great iron and steel industries centering in Birmingham, Alabama.

SOUTHEASTERN COASTAL PLAIN

A distinctive coastal plain margins North America, along the coastal zone of the Atlantic Ocean and the Gulf of Mexico, in varying widths but nearly unbroken from Cape Cod and Long Island to Yucatan in southeastern Mexico. In most parts, very flat seaward margins of low-lying and most recent sedimentary deposits are followed inland by a tendency for a coastwise banding of features. A landward increase in elevation and erosional dissection also tends to contribute toward rolling and irregular plains with local relief up to a very few hundreds of feet A gradual gentle uplift or upwarp of the geologically recent sedimentary rock layers means that they are little disturbed from their original horizontal attitude and dip slightly seaward and that such regions have most uplift inland toward bordering backland areas of differing character Erosion, therefore, bares lower and somewhat older sedimentary layers progressively in a landward direction Landward-facing cuestas or bands of low hills may develop on more-resistant rock exposures; whereas lower relief and occasional flat intervales develop by differential erosion in exposures of less-resistant rock layers. The farthest inland of these, commonly developed in Cretaceous clays, are termed inner lowland (see ESCARPMENT). The landward coastal plain boundary is commonly the inner margin of the Cretaceous clays, and the border between the coastal plain layers and the oldland or backland rocks is known as the fall line or fall zone. See FALL LINE.

Several other factors influence differences within the great extent of these southeastern coastal plains. Recent changes in relative elevation of land and sea constitute such an influence A predominance of land subsidence marks the region from North Carolina to Cape Cod with many embayments and shorelines irregular in contrast to most other parts, of recent upraisal or stillstand of the land relative to the sea. A strongly marked variance from north to south in climatic characteristics with an especially steep temperature gradient is reflected in such aspects as natural vegetation (Fig. 4) and length of growing season and agricultural production, as well as the pattern contrasts in winter resorts of Florida and the Gulf coast versus the definite winter influences of the New York and New Jersey coastal zones. See AIR TEMPERATURE; SOIL; SOIL, ZONAL DISTRIBUTION; VEGETATION ZONES (WORLD). Differing widths of valley bottom alluvium appear along streams and are a large and significant factor in case of valleys of trunk streams following the regional slope from the backland to the sea—the Apalachicola, Rio Grande, and particularly the great valley and delta of the Mississippi. See COASTAL PLAIN; DELTA; FLOOD PLAINS; FLUVIAL EROSION CYCLE; FLUVIAL EROSION LANDFORMS.

Within the essential unity of this great coastal plain form and structure, six or seven subdivisions are useful in outlining the regional variations characterizing its long extent: Northern embayed, Sea Island, Peninsular Florida, East Gulf section, Mississippi alluvial plain, West Gulf, and Yucatan section.

Northern embayed section. Considerable evidence indicates that the coastal plain features northward from Cape Fear and the estuary of the Cape Fear River, although affected by both emergence and submergence, are most marked from a recent predominance of submergence in relation to tidewater The outer coastal plain flats are not only marked by swamps and marshes but tidewater extends upvalley inland in the North Carolina sounds and in the drowned-valley estuaries of Chesapeake and Delaware Bays. Bars and barrier beaches form on the longshore margin of the sediments extending beneath the shallow sea, partially enclosing the lagoons and other tidewater features of this coastal plain section. See COASTAL LANDFORMS; MARINE MARSH; SHORE PROCESSES

From Long Island to Cape Cod, subsidence makes most of the recent coastal plain, including the inner lowland (as in Long Island Sound), now a part of the continental shelf. This portion was glaciated, and moraines augment the cuestas and erosional remnants of the coastal plain that remain as islands and peninsulas, Long I... Martha's Vineyard, Nantuck...

Proximity to great urban areas is reflected from Long Island through New Jersey and eastern Maryland Truck farming utilizes many areas of sandy soils, and the beaches include some of the most outstanding coastal resorts in the United States Much of the more southern portions is returning to forest uses along with continuing truck farming and specialized agriculture Along with scattered urban population, some of the great urban areas center upon the older seaports, such as New York and Baltimore at the fall line

Sea Island section. The Coastal Plain in South Carolina and Georgia is broad and comparatively simple in pattern The fall zone is a nearly indistinguishable transition marked by low hills and several cities, such as Columbus, Macon, and Augusta, Georgia; and Columbia, South Carolina Regrowing forests, grazing, and agriculture including early-season truck farming utilize much of this area Wet lands and tidewater intermingle in swamp forest, marsh, and muddy or sandy islands in the outer Coastal Plain The name Sea Island is taken from the repeated coastal pattern of longshore islands The outer lowland plain contains a few urban foci at seaport locations such as Charleston, Savannah and Brunswick.

Peninsular Florida. This low relief peninsula results from a recent anticlinal upwarping of a por-[illegible]

[illegible] depressions, underground drainage, and great springs, developed in the limestone rock Marine organic sediments yield valuable phosphate materials Much of the rural land of the north is taken up by piney forest, considerable grazing, and citrus fruit raising A great deal of the land of southern Florida is very low, youthfully exposed plains surface Lake Okeechobee and the poorly drained Everglades are examples of initially undrained depressions in such a land surface, but artificial drainage projects are reclaiming more land for fruit and truck farming for off-season and other marketing in northern United States

Tropical mild winters and attractive coastal zone patterns stimulate a variety of developments Jacksonville and Miami are examples of growing seaports and industrial places as well as resort centers Older shell beach ridges, lagoons, and current barrier beaches comprise much of the eastern shore with its speculative and resort developments This coastal line, partly barrier bar and partly coral in origin, swings southward in the chain of the Keys to the marine base at Key West Offshore to the southeast a broad protrusion of continental shelf is marked by the Bahama Islands in a zone of [illegible] rida

[illegible] sections. The Mississippi alluvial plain divides these two coastal plain segments bordering the Gulf of Mexico. The eastern part, in Alabama and Mississippi, presents the typical land features of the belted coastal plain with cuestaform ridges separating intervale and inner lowlands. The previously rich agricultural land of the inner lowland in the Alabama Black Belt is faced with problems of depletion of soil fertility under present plans of use and from competition with cultivation farther west and in the interior plains. As well as continuing agriculture, these coastal plain lands support grazing in forest and grassy openings, and the piney woods are regrowing as sources of lumber and turpentine. Some petroleum fields are producing from local domes, folds, and entrapping structures within the coastal plain rocks. Small cities and towns are mostly scattered [illegible] bama.

The West Gulf Coastal Plain continues from the Mississippi lowlands to the Rio Grande as a wide belted coastal plain but with certain marked differences. The fertile inner lowland terminates against an abrupt fault scarp, most of which is the Balcones Escarpment [illegible] by the cities of Dallas, Fort Worth, Austin, and San Antonio. The outer Coastal Plain, with barrier beaches, lagoons, estuarial bays, and marshes, maintains channeled seaport contact through Galveston, Houston, Beaumont, Port Arthur, and Corpus Christi, which are also growing industrial places Notable spots of subtropical agriculture now mark the coastal zone in machine cultivation of rice, citrus and other fruits, and truck raising, especially citrus in the lower Rio Grande vicinity.

Petroleum fields have been prospected with geophysical and other techniques to develop some of [illegible] ments off shore in the continental shelf This happened first in California, then in the Gulf of Mexico, and is recently showing promise of future success in the Bahama section and possibly in the Atlantic continental shelf off the eastern United States

The Gulf Coastal Plain narrows southward in Mexico, is interrupted in several places, and becomes too hot and steaming a tropical lowland in Mexico to attract much settlement. Tampico and Veracruz are seaport cities with rail and highway connection into highland Mexico Petroleum production declined in this area, but it is now being revived on a considerable scale The volcanic highlands of Mexico end in abrupt slopes to the sea northward of Veracruz, and the striking Tuxtlas volcanoes rise on the seaward margin of the coastal plain to the south of that city Beyond, at the Isthmus of Tehuantepec, the coastal plains of the Gulf of Mexico and of the Pacific Gulf of Tehuantepec are nearly joined.

Yucatan section. More coral than limestone rock material, more definitely tropical climate and

some extent. Petroleum extraction continues in several parts. The central portion is in the heart of the Appalachian bituminous coal fields and coal industries. The great urban center of Pittsburgh stands within the region at the head of the Ohio Valley. There coal, coke from neighboring industries, and various iron ores have been combined to make this a leading iron and steel manufacturing center. Other chief centers of population in the valleys and a few such as Charleston and Wheeling, West Virginia, have become sizable urban centers.

5. Cumberland section. In this southern portion the plateau belt narrows in width and gradually decreases in local relief. The breadth of the generally slight synclinal structure is broken by a few longitudinal folds These anticlinal parts are etched out as valleys between large, long, flat-topped remnants of the plateau upland, such as Sand Mountain and West Sand Mountain. Lookout Mountain is a similar remnant, judged by some to belong to the Folded Appalachians and by others to be part of the plateau system. The Tennessee River cuts an irregular course across the region and forms many valley sections now developed by dams and reservoirs as important links in the TVA system. Farther north the eastern part of the plateau, called the Cumberland Mountains, contains historic Cumberland Gap. This is actually a narrow valley ascending the eastern margin of the upland near the border junction of Virginia, Kentucky, and Tennessee, and is a route still used by railroads and highways.

Fine coal fields have been developed in the Kentucky, Tennessee, and Alabama parts of the plateau region. The Warrior coal field contributes largely to the great iron and steel industries centering in Birmingham, Alabama

SOUTHEASTERN COASTAL PLAIN

A distinctive coastal plain margins North America, along the coastal zone of the Atlantic Ocean and the Gulf of Mexico, in varying widths but nearly unbroken from Cape Cod and Long Island to Yucatan in southeastern Mexico In most parts, very flat seaward margins of low-lying and most recent sedimentary deposits are followed inland by a tendency for a coastwise banding of features. A landward increase in elevation and erosional dissection also tends to contribute toward rolling and irregular plains with local relief up to a very few hundreds of feet. A gradual gentle uplift or upwarp of the geologically recent sedimentary rock layers means that they are little disturbed from their original horizontal attitude and dip slightly seaward and that such regions have most uplift inland toward bordering backland areas of differing character. Erosion, therefore, bares lower and somewhat older sedimentary layers progressively in a landward direction. Landward-facing cuestas or bands of low hills may develop on more-resistant rock exposures; whereas lower relief and occasional flat intervales develop by differential erosion in exposures of less-resistant rock layers. The farthest inland of these, commonly developed in Cretaceous clays, are termed inner lowland (*see* ESCARPMENT). The landward coastal plain boundary is commonly the inner margin of the Cretaceous clays, and the border between the coastal plain layers and the oldland or backland rocks is known as the fall line or fall zone. *See* FALL LINE.

Several other factors influence differences within the great extent of these southeastern coastal plains. Recent changes in relative elevation of land and sea constitute such an influence. A predominance of land subsidence marks the region from North Carolina to Cape Cod with many embayments and shorelines irregular in contrast to most other parts, of recent upraisal or stillstand of the land relative to the sea. A strongly marked variance from north to south in climatic characteristics with an especially steep temperature gradient is reflected in such aspects as natural vegetation (Fig. 4) and length of growing season and agricultural production, as well as the pattern contrasts in winter resorts of Florida and the Gulf coast versus the definite winter influences of the New York and New Jersey coastal zones. *See* AIR TEMPERATURE; SOIL; SOIL, ZONAL DISTRIBUTION; VEGETATION ZONES (WORLD) Differing widths of valley bottom alluvium appear along streams and are a large and significant factor in case of valleys of trunk streams following the regional slope from the backland to the sea—the Apalachicola, Rio Grande, and particularly the great valley and delta of the Mississippi *See* COASTAL PLAIN; DELTA; FLOOD PLAINS; FLUVIAL EROSION CYCLE; FLUVIAL EROSION LANDFORMS

Within the essential unity of this great coastal plain form and structure, six or seven subdivisions are useful in outlining the regional variations characterizing its long extent: Northern embayed, Sea Island, Peninsular Florida, East Gulf section, Mississippi alluvial plain, West Gulf, and Yucatan section.

Northern embayed section. Considerable evidence indicates that the coastal plain features northward from Cape Fear and the estuary of the Cape Fear River, although affected by both emergence and submergence, are most marked from a recent predominance of submergence in relation to tidewater. The outer coastal plain flats are not only marked by swamps and marshes but tidewater extends upvalley inland in the North Carolina sounds and in the drowned-valley estuaries of Chesapeake and Delaware Bays. Bars and barrier beaches form on the longshore margin of the sediments extending beneath the shallow sea, partially enclosing the lagoons and other tidewater features of this coastal plain section. *See* COASTAL LANDFORMS; MARINE MARSH; SHORE PROCESSES.

From Long Island to Cape Cod, subsidence makes most of the recent coastal plain, including the inner lowland (as in Long Island Sound), now a part of the continental shelf. This portion was glaciated, and moraines augment the cuestas and erosional remnants of the coastal plain that remain as islands and peninsulas, su[illegible] Long I[illegible] Martha's Vineyard, Nantuc[illegible]

Proximity to great urban areas is reflected from Long Island through New Jersey and eastern Maryland Truck farming utilizes many areas of sandy soils, and the beaches include some of the most outstanding coastal resorts in the United States Much of the more southern portions is returning to forest uses along with continuing truck farming and specialized agriculture Along with scattered urban population, some of the great urban areas center upon the older seaports, such as New York and Baltimore at the fall line

Sea Island section. The Coastal Plain in South Carolina and Georgia is broad and comparatively simple in pattern The fall zone is a nearly indistinguishable transition marked by low hills and several cities, such as Columbus, Macon, and Augusta, Georgia; and Columbia, South Carolina Regrowing forests, grazing, and agriculture including early-season truck farming utilize much of this area Wet lands and tidewater intermingle in swamp forest, marsh, and muddy or sandy islands in the outer Coastal Plain The name Sea Island is taken from the repeated coastal pattern of longshore islands The outer lowland plain contains a few urban foci at seaport locations such as Charleston [illegible]

[illegible]

Coastal Plain is marked in the north by numerous karst features, such as sinkholes, shallow lake-filled depressions, underground drainage, and great springs, developed in the limestone rock. Marine organic sediments yield valuable phosphate materials Much of the rural land of the north is taken up by piney forest, considerable grazing, and citrus fruit raising A great deal of the land of southern Florida is very low, youthfully exposed plains surface. Lake Okeechobee and the poorly drained Everglades are examples of initially undrained depressions in such a land surface, but artificial drainage projects are reclaiming more land for fruit and truck farming for off-season and other marketing in northern United States

Tropical mild winters and attractive coastal-zone patterns stimulate a variety of developments. Jacksonville and Miami are examples of growing seaports and industrial places as well as resort centers Older shell beach ridges lagoons, and current barrier beaches comprise much of the eastern shore with its speculative and resort developments This coastal line, partly barrier bar and partly coral in origin swings southward in the chain of the Keys to the marine base at Key West Offshore to the southeast a broad protrusion of [illegible]

[illegible] these two coastal plain segments bordering the Gulf of Mexico The eastern part, in Alabama and Mississippi, presents the typical land features of the belted coastal plain with [illegible] separating intervale and inner lowlands. The previously rich agricultural land of the inner lowland in the Alabama Black Belt is faced with problems of depletion of soil fertility under present plans of use and from competition with cultivation farther west and in the interior plains As well as continuing agriculture, these coastal plain lands support grazing in forest and grassy openings, and the piney woods are regrowing as sources of lumber and turpentine. Some petroleum fields are producing from local domes, folds, and entrapping structures within the coastal plain rocks. Small cities and towns are mostly scattered in the inland lowland belts except for a few in the outer coastal plain, such as Tallahassee and Pensacola in Florida and Mobile, Alabama

The West Gulf Coastal Plain continues from the Mississippi lowlands to the Rio Grande as a wide belted coastal plain but with certain marked differences The fertile inner lowland terminates against an abrupt fault scarp, most of which is the Balcones Escarpment up to the high plains of western Texas. Eastward the Coastal Plain is bordered by the Ouachita upland The inner belt is marked by the cities of Dallas, Fort Worth, Austin, and San Antonio The outer Coastal Plain, with barrier beaches, lagoons, estuarial bays, and marshes, maintains channeled [illegible] Galveston, Houst[illegible] Corpus Christ[illegible] places. Notable spots of subtropical agriculture now mark the coastal zone in machine cultivation of rice, citrus and other fruits, and truck raising, especially citrus in the lower Rio Grande vicinity.

Petroleum fields have been prospected with geophysical and other techniques to develop some of the continent's outstanding gas and oil productions [illegible]

[illegible] petroleum exploitation into the submarine sediments offshore in the continental shelf. This happened first in California, then in the Gulf of Mexico, and is recently showing promise of future success in the Bahama section and possibly in the Atlantic continental shelf off the eastern United States.

The Gulf Coastal Plain narrows southward in Mexico, is interrupted in several places, and becomes too hot and steaming a tropical lowland in Mexico to attract much settlement. Tampico and Veracruz are seaport cities with rail and highway connection into highland Mexico Petroleum production declined in this area, but it is now being revived on a considerable scale. The volcanic highlands of Mexico end in abrupt slopes to the sea northward of Veracruz, and the striking Tuxtlas volcanoes rise on the seaward margin of the coastal plain to the south of that city. Beyond, at the Isthmus of Tehuantepec, the coastal plains of the Gulf of Mexico and of the Pacific Gulf of Tehuantepec are nearly joined

Yucatan section. More coral than limestone rock material, more definitely tropical climate and veg

etation, and less settlement and development modify an otherwise striking comparison between the peninsulas of Yucatan and Florida. Such karstland features as many sinkholes and underground drainage are widespread. Mayan Indian civilization has left many remains of scattered towns of the past. Today this peninsular plain has areas of agricultural development in sugar cane, maize, tobacco, coffee, and the plantation production of sisal fiber. A large part of the region, however, remains a wilderness of tropical scrub land on the north and tropical forest inland to the south

Mississippi alluvial plain. Beginning at about Cairo, Ill., the Mississippi River swings in an immensely meandering course first on one side and then on the other of a broad flood plain (50–100 mi wide) for 600 mi before discharging over a birdsfoot delta into the Gulf of Mexico. The junctions of tributaries are dragged downstream, with the migration of swinging meanders, so that they commonly flow downvalley for great distances before actually joining the main stream—as is the case of the Yazoo River and Basin. The detailed landforms of streamside natural levees, meander scars, oxbow lakes, and poorly drained back marshes are typical of the old age stage in the development of a river valley (*see* PLAINS). These features appear in modified but characteristic patterns in the great delta.

Several physical attributes make this plain both valuable and troublesome for human use. The soils are fertile and the climate largely humid subtropical. The river ways have been used for travel and commerce from the beginning of European settlement to the present, but always in the face of the difficulties of flooding, silting, and shifting courses Earlier attempts to control the river by enlarging the natural levees and by strict confinement met with recurrent disaster. Today somewhat better results are attained by a combination of drainways and flooding areas along with river confinement in strategic places.

A few smaller cities and one great one are economic and cultural foci of the river plain. Baton Rouge, Natchez, Vicksburg, and Memphis are examples of the smaller urban centers. New Orleans, on the delta, continues to dominate the lower Mississippi and is its seaport connection with the Gulf, Caribbean, South America, and more distant places.

ARCTIC MARGIN PLAINS

Two plains units that are commonly considered parts of the interior North American lowlands are here separately considered because of their high latitude location and features Both—the Arctic slope of Alaska and the Mackenzie plains of Canada—open out to Arctic Ocean shores from which the polar ice pack recedes for an unpredictable duration during the late part of the warm season Ships do touch these Arctic shores briefly, and both areas are on the great circle routes by air between the eastern United States and the Far East or much of the Asiatic U.S.S.R. Both are now served by air for local and external contacts. These plains are subject to Arctic cold and long hours of dark ness during the low-sun period, and conversely to long hours of daylight with short but surprisingly warm periods during the high sun. Both are strongly marked in their lower parts by small lakes and ponds, polygonal ground, and other features of permanently frozen ground (see PERMAFROST) They are also subject to the widespread flooding of lower and flatter parts which characterizes subarctic regions, especially when the headwaters thaw earlier than the ice breakup in the main streams to the northward.

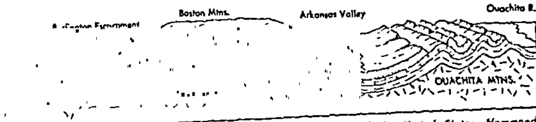
Alaskan Arctic slope. Variable widths of this sloping treeless tundra extend northward from the Brooks Range to Arctic shores. Most of the middle and outer parts have true coastal plain characteristics, but the inner portion varies from rolling foothill plateau at about 2000 ft elevation to a hilly upland, in the western part, strongly marked by partial erosion of domes and folds. The latter results in a pattern of hills like those in the Newer or Folded Appalachians. Much of this slope region has been a U.S. Naval petroleum reserve, and oil has been recovered from reservoir structures near Barrow on the coast and Umiat farther in the center. A few Eskimos and research posts occupy the area, which is also invaded during summer by migrating herds of caribou and other herbivores and their itinerant hunters The brief Arctic summer brings out great numbers of flowers, migrating birds, myriads of insects, and numerous small animals.

Mackenzie plains. Cuestaform plains, like those of the interior, here extend to the Arctic shores. The Mackenzie River drains most of this subarctic extension of the interior cuestaform plains between the Canadian Shield and the Rocky Mountain cordillera The lower course is in an alluvial plain and delta of poorly drained, watery, permafrost and tundra country that is easier to cross when frozen, but it is contacted by air and by summer navigation of river and lake steamers The boreal forest or taiga, progressively sparser and more stunted in height to the north, covers much of the land except for marsh and muskeg of the lower parts. Hunting and trapping for furs continue, but agriculture is scarcely feasible by present practices farther north than the Peace River country. Scattered mines for radium, uranium, and other metallic minerals in the bordering rocks of the Shield are linked to outside markets by air and by the river and lakes routes (such as Great Slave and Great Bear Lakes) on the borders between the Shield and Mackenzie regions.

These plains are known to contain locally entrapped reservoirs of petroleum, as at Norman Wells, but they remain a future reserve in the face of competition with those closer to markets, as in the interior plains.

INTERIOR PROVINCES

Interior North America is so predominantly plainsland, although it contains a few highland parts, that other physical · onl



19 Structure-topography diagram of the Ozark Ouachita provinces (From A K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

e as much to regional differences as do structure and surface form of the plains The aspects of area and huge distances often cause the extent and character of climate and its closely associated problems of natural vegetation (Fig 4) soil to make the greatest regional physical contrasts in many parts of the plains. See CLIMATOLOGY; SOIL, ZONAL DISTRIBUTION; VEGETATION ZONES (WORLD) Recent (Pleistocene) glaciation north of the Ohio and Missouri river valleys leaves its impress on patterns of drainage, surface forms, and even soil materials quite different from areas not recently affected by glaciers (see GLACIATED TERRANE) The map and list of Fig. 2 indicate the name and location of the five provinces and their principal sectional divisions of the great interior of the continent The predominant characteristics of each division are outlined briefly below For illustration and detailed considerations, *see* EARTH RESOURCE PATTERNS; TERRAIN AREAS, WORLD-WIDE

Two other matters of general character and regional differences are perhaps best outlined for the whole interior of the continent Only a few parts are extensive flat plains, so featureless as to be conspicuous Most of these are exposed bottoms of larger lakes of the glacial period such as the Lake Maumee Plain (west of Lake Erie) the Lake Agassiz Plain (Minnesota North Dakota, and southern Manitoba) and the area to the west of Great Slave Lake Nearly all of the others are areas of some local relief and are perhaps best termed low relief of rolling and irregular plains. The little disturbed sedimentary rocks are seldom perfectly horizontal but are repeatedly bowed or warped into domed or basin shapes Erosion tends to degrade the higher and weaker parts first, so that rings of hills or infacing cuestaform ridges tend to develop around an erosional basin cut into a domed structure Conversely outfacing cuestas and rings of hills tend to develop around a structural basin while erosion is beveling off at differential rates the more- and less-resistant rock layers on the higher margins of the basin structure. Scattered places develop enough local relief to be classed as hilly country and parts of the high Great Plains are here and there cut into tablelands by scattered but deep stream valley dissection Only the Ouachita Mountains and parts of the Black Hills (Fig. 12) in the Great Plains province, have some low mountain terrain.

Interior low plateaus. Westward from the Appalachian plateau in Kentucky and Tennessee slopes a gently up arched limestone surface pocked by sinkholes and solution caverns This upland, called the Highland Rim, contains two domed parts now being eroded into the Kentucky Bluegrass or Lexington Plain and, to the southward, the Nashville Dome, now presenting the eroded basin of the Nashville Basin or Tennessee Bluegrass The cuestas and eroded margins of the Highland Rim make hilly uplands surrounding the two local plains. The local plains are regions of agriculture, animal raising, and some urban development.

Ozark province. This is a hilly upland in a slightly disturbed, locally warped and scarped domelike plateau Lead is mined near Joplin on the western margin, and some iron is extracted and used from granitic rocks exposed in the eroded local domes of the so-called St. Francis Mountains on the east. The *Springfield-Salem plateau section* is varied hilly upland cut into an upraised and scarped structure Springfield is a small city and cross-route center in the irregularly settled plateau. The Boston Mountains, a dissected portion of the plateau structure, have the roughest and wildest portions for some 200 mi east-west on the southern margin of the Ozarks

Ouachita province. One broad anticlinal structure and two large synclinal structures in the Paleozoic rock layers are variously affected by erosion to cause extension of this highland province 200 mi westward from the Mississippi alluvial plain through central Arkansas into Oklahoma

Between the Boston Mountains on the north and the Ouachita Mountains on the south the locally wrinkled syncline or synclinorium of the Arkansas Valley is a structural depression that also contains the swampy alluvial plain of the meandering Arkansas River Much of this lowland has been used for cotton raising More-resistant sandstones of the wrinkled rock layers remain upstanding as scattered linear ridges

South of the Arkansas Valley the anticlinorium is erosively etched by streams in a trellis pattern into hill and low mountain country resembling that of the Folded or Newer Appalachians. Local relief ranges up to 1500 ft, and the more rugged portions are largely wild and forested land Hot Springs National Park is found in the southeast, where there are several major faults with dykes and more massive intrusions of igneous rocks Most of the southern margins of this mountain section are in the second synclinorial zone, where the previously beveled (peneplaned) rock layers

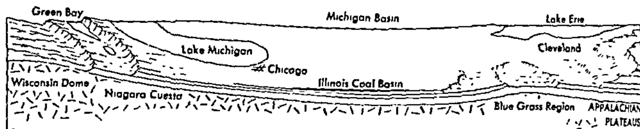


Fig. 11. Structure and surface diagram of a part of the interior lowlands. (*From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957*)

in elevation and are overlapped by the Gulf Coastal Plain *See* FALL LINE.

Central Lowland province. A great three-pronged area of interior plainslands spreads between the more eastern highlands of the Laurentian or Canadian Shield, the Appalachian Highlands, and the Ozark-Ouachita provinces, and the western higher plains in the Great Plains province. The geographical heart of this low-relief interior, which led in the settlement and development of the Continent, is commonly called the Midwest and Great Lakes regions of the United States and Canada These regions coincide fairly closely with the center and eastern prong of the great Central Lowland reaching from the Corn Belt agricultural region eastward through the Great Lakes country into the Mohawk depression toward the Hudson River Valley. All of this center and eastern prong except the Wisconsin Driftless section has been recently affected by Pleistocene glaciation, as has the large northwestern prong, the Western Lake section. (For the character of surface features of plains, modified by recent glaciation, *see* GLACIATED TERRANE; PLAINS.) Only a third prong, the smaller southwestern protrusion called the Osage section, has not been so modified.

Eastern Lake Section. Here great lobes of the glaciers weighed down and enlarged depressions of earlier terrain to leave, after they melted, all the Great Lakes except Superior These waterways and the lowland route to tidewater through the Hudson-Mohawk depression have long been important in the settlement and development of the continent, but their value is augmented by the completion of the Great Lakes–St. Lawrence Seaway. Iron ore, from Superior fields of the Shield and from other places, has been readily brought to meet with coal and coke from Appalachian and Illinois fields Industrial and commercial growth of cities and towns has gone hand in hand with the development of waterways, railroads, and roadways in a net commonly focusing on or funneling through this region. Perhaps the greatest urban focus developed at Chicago on the south end of Lake Michigan, but other cities, such as Milwaukee, Detroit, Cleveland, and many smaller ones, have grown in the area. Originally a great lumbering region, its regrowing forests, many inland lakes, and park systems have stimulated growth of tourism and resorts. With a variety of sandy to loamy soils, the area is used for hay and dairying, general farming, and specialized fruit growing (just east of Lake Michigan, along the south shores of Lake Erie, and in the Niagara country between Lakes Ontario and Erie).

The broad shallow structural basin centering on the southern peninsula of Michigan is rimmed by the Niagara escarpment. The basin contains the Michigan petroleum fields and some coal The margining escarpment is conspicuous in Wisconsin where it nearly separates Green Bay from Lake Michigan, and in Canada where it forms much of the separation of Georgian Bay from Lake Huron The escarpment stands out plainly between Lake Erie and Lake Ontario, and the Niagara River falls spectacularly into the gorge it has notched through the resistant Niagara limestone.

Wisconsin Driftless area. Between the Eastern and Western Lake sections this area was unmodified by Pleistocene glaciation. It lies mostly in Wisconsin and serves as an example of the Midwestern interior before glacial alteration. Low-relief hills and valleys and other forms result from fluvial erosion. Although it contains some lead and zinc deposits, the Driftless area is mostly an agricultural region, although the soils are less desirable than in several adjoining areas.

Western Lake section. This low-relief plains section is covered by glacial drift and dotted by lakes, large and small It lies on the transition between the originally forested lands with brown or gray soils and the prairie grasslands with their deep staining of humus and great natural fertility. Winters are long and severe, while summers are continentally warm to hot Spring-wheat raising has been extended from Iowa through Minnesota and the Dakotas northward into the Canadian frontier toward the narrow junction with the Mackenzie Lowland section.

Till plains and dissected till plains sections These two sections coincide closely with the fertile American Corn Belt. The eastern or till plains section was heavily mantled with glacial till during the later stages of glaciation Originally forested, with a fertile brown forest soil, it was early cleared and valued for cultivation of corn (maize) and forage to produce beef and pork. The Illinois part contains the coal basin The region is now dotted with industrial towns in Illinois, Indiana, and Ohio

The dissected till section west of the Mississippi is a plainsland of older drift, completely dissected by streams in valleys cutting as much as a few hundred feet of local relief Since this was also the

rairie grassland area, fertile drift materials were ven more darkly and richly stained This is also neat-raising farm country with packing, trading, nd some industry

Osage section. Because this is the western slope of the Ozark uplift, the Paleozoic rock layers dip westward These layers are somewhat beveled by erosion and there are banded north-south rougher hilly lands margining, like east-facing cuestas, the plateaulike harder-rock uplands. The dipping structure of this southwesterly prong of the Central Lowland province is overlapped by later (Cretaceous) rocks of the Great Plains on the west Cultivation of corn on the north and cotton in the south is the chief agricultural use of this land near the dry margins of such types of farming

to

lands to the west, underlie these plains with increasing surface elevations from east to west. Geologically the plains are a great shallow synclinal basin elongated from north to south. In many places, petroleum entrapped in local parts of the

and the great distance to major markets generally preclude mining development. Many of the traditional subdivisions of the region (*see map of* Fig. 2) are based on progress and depth of erosion, recent glaciation or its lack of influence in the land character, local doming or upfolding, structural intrusion or volcanism, and near-plateau character of many parts of these elevated plains

Most of the Great Plains was originally grassland, from short prairie to semiarid steppe land in the dry continental interior Dry-margin winter-wheat raising

spring-whe

plains to th

the Rockies in the Saskatchewan plateau section (Fig 2) Recently this spring-wheat raising has been successfully extended, along with cattle raising, as far north as the Peace River plains at the northern margin of the Great Plains province Most other parts have limited farming and are more extensively utilized for ranging and grazing of cattle and sheep Some cotton growing is, however, established in the southern margins of the province. See MINERAL FUEL AREAS; PRAIRIE, STEPPE

Because of the harshness of the continental interior and cordilleran blocking of Pacific air masses, these continental high plains remain, in general, some of the least populated and least developed parts of the continent, with a few scattered exceptions.

CORDILLERAN NORTH AMERICA

A large proportion of western North America is a great north-south highland of mountain masses, ranges and systems, and of elevated basins and plateaus. With the Central American ranges and mountains of somewhat different structure and axial orientation, the North American Cordillera continues from those of Andean South America, the great mountain rim of the Pacific. These Pacific-encircling mountains turn westward in Alaska and swing toward the Asian shores of the Pacific in the great arcuate festoons of the Aleutian, Kuril, and Japanese islands.

The consequences in the physical and human geography of the continent are pronounced but somewhat in contrast with those of South America. The great orographic barrier made difficult an articulation of settlement and development between the eastern regions and the western highlands and coastal zones. This difficulty was only resolved by the penetration of railways and telecommunications through the barrier passes, shortened water routes via Panama Canal, and a later extension of motor highways and aviation routes. The greatest contrasts appear, however, in the patterns of climate and reflected vegetation because most, and the broader parts, of North America lie in the middle and higher middle latitudes. In these zones of prevailing west-to-east air-mass movements, the north-south Cordillera blocks easy access of the moderate and moisture-bearing Pacific air masses. Thus much of the western highland areas and a large proportion of the great interior plains have extremes of continental temperatures and meagerness of precipitation (*see desert, steppe, and prairie vegetation of* Fig 4), in contrast to the marine moderation and moisture which extend far into European plains *See* CONTINENTALITY, WEATHER AND CLIMATE, MARINE INFLUENCE ON WEATHER AND CLIMATE.

Although bold in plan and extent, the Cordilleran lands are diverse in component highland parts and commonly intricate in local details of

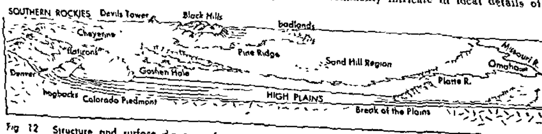


Fig 12 Structure and surface diagram of part of the Great Plains (From A K Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

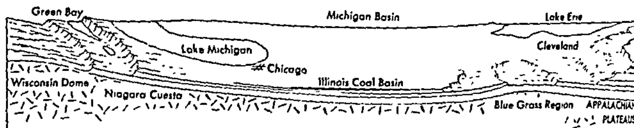


Fig 11. Structure and surface diagram of a part of the interior lowlands. (From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

in elevation and are overlapped by the Gulf Coastal Plain. *See* FALL LINE.

Central Lowland province. A great three-pronged area of interior plainslands spreads between the more eastern highlands of the Laurentian or Canadian Shield, the Appalachian Highlands, and the Ozark-Ouachita provinces, and the western higher plains in the Great Plains province. The geographical heart of this low-relief interior, which led in the settlement and development of the Continent, is commonly called the Midwest and Great Lakes regions of the United States and Canada. These regions coincide fairly closely with the center and eastern prong of the great Central Lowland reaching from the Corn Belt agricultural region eastward through the Great Lakes country into the Mohawk depression toward the Hudson River Valley. All of this center and eastern prong except the Wisconsin Driftless section has been recently affected by Pleistocene glaciation, as has the large northwestern prong, the Western Lake section. (*For the character of surface features of plains, modified by recent glaciation, see* GLACIATED TERRANE, PLAINS) Only a third prong, the smaller southwestern protrusion called the Osage section, has not been so modified.

Eastern Lake Section. Here great lobes of the glaciers weighed down and enlarged depressions of earlier terrain to leave, after they melted, all the Great Lakes except Superior These waterways and the lowland route to tidewater through the Hudson-Mohawk depression have long been important in the settlement and development of the continent, but their value is augmented by the completion of the Great Lakes–St Lawrence Seaway. Iron ore, from Superior fields of the Shield and from other places, has been readily brought to meet with coal and coke from Appalachian and Illinois fields. In- [illegible] haps the greatest urban focus developed at Chicago on the south end of Lake Michigan, but other cities, such as Milwaukee, Detroit, Cleveland, and many smaller ones, have grown in the area Originally a great lumbering region, its regrowing forests, many inland lakes, and park systems have stimulated growth of tourism and resorts. With a variety of sandy to loamy soils, the area is used for hay and dairying, general farming, and specialized fruit growing (just east of Lake Michigan, along the south shores of Lake Erie, and in the Niagara country between Lakes Ontario and Erie)

The broad shallow structural basin centering on the southern peninsula of Michigan is rimmed by the Niagara escarpment. The basin contains the Michigan petroleum fields and some coal. The margining escarpment is conspicuous in Wisconsin where it nearly separates Green Bay from Lake Michigan, and in Canada where it forms much of the separation of Georgian Bay from Lake Huron. The escarpment stands out plainly between Lake Erie and Lake Ontario, and the Niagara River falls spectacularly into the gorge it has notched through the resistant Niagara limestone.

Wisconsin Driftless area. Between the Eastern and Western Lake sections this area was unmodified by Pleistocene glaciation. It lies mostly in Wisconsin and serves as an example of the Midwestern interior before glacial alteration. Low-relief hills and valleys and other forms result from fluvial erosion. Although it contains some lead and zinc deposits, the Driftless area is mostly an agricultural region, although the soils are less desirable than in several adjoining areas

Western Lake section. This low-relief plains section is covered by glacial drift and dotted by lakes large and small. It lies on the transition between the originally forested lands with brown or gray soils and the prairie grasslands with their deep staining of humus and great natural fertility. Winters are long and severe, while summers are continentally warm to hot. Spring-wheat raising has been extended from Iowa through Minnesota and the Dakotas northward into the Canadian frontier toward the narrow junction with the Mackenzie Lowland section.

Till plains and dissected till plains sections These two sections coincide closely with the fertile American Corn Belt. The eastern or till plains section was heavily mantled with glacial till during the later stages of glaciation. Originally forested, with a fertile brown forest soil, it was early cleared and valued for cultivation of corn (maize) and forage to produce beef and pork The Illinois part contains the coal basin The region is now dotted with industrial towns in Illinois, Indiana and Ohio

The dissected till section west of the Mississippi is a plainsland of older drift, completely dissected by streams in valleys cutting as much as a few hundred feet of local relief. Since this was also the

partake of the character of bordering regions as well as of those of the Southern Rocky Mountains

Distinctive highlands separate, on the south and west, the Wyoming Basins from the Colorado Plateaus and the Great Basin near the Great Salt Lake. West of Cheyenne and on the other side of the Southern Rockies the Uinta Mountains extend east west through northern Utah to adjoin the orth-south Wasatch Mountains just east of Salt ake City and near the southwestern corner of Vyoming This roughly eroded upwarp of the sedimentary layer of the Colorado Plateau structures nakes a high, rugged barrier on the south which is harply breached by the Green River as it passes hrough the Flamina Gorge and, after an eastward ffset, cuts south through narrow Ladore Canyon to become a major tributary of the Colorado River. The lofty Wasatch Mountains, a north-south block-mountain range between Great Salt Lake and the Wyoming Basin, are continued northward in a slightly east-bowing arc as a series of parallel block-mountain ranges, like those of the Great Basin but more closely spaced These then continue curving slightly west to pass by the south end of the Teton Range and disappear beneath the lava layer of the Snake River Plateau, one part of the Columbia Plateau system.

The Teton Range (now included in the Grand Teton National Park) is a short but majestic range, deeply cut by erosion but with its summits reaching above 12,000 ft, several thousand feet above timber line and with many snow-covered peaks The whole range overtowers the Jackson Hole, one of the small, nearly enclosed basins of the Wyoming Basin region

The Yellowstone section includes the elevated faulted, and occasionally deeply eroded basin plateau, which is the heart of Yellowstone National Park, and its irregularly bordering mountains The mountains, named in a clockwise direction, are the Madison and Gallatin Ranges, with their ends in the northwest of the Park, the Snowy (on the north), and the Absaroka Ranges (on the east) The Tetons extend south from the southwestern area of the park Deeply gashed canyons the basin lake (Yellowstone), and the great scarps with many associated hot springs and geysers are some of the spectacular features of Yellowstone National Park

East of Yell... enclosed by ... Owl Creek ... onto the G... ...ains on the north, but is notably drained by rivers tributary to the Big Horn, which flows in through the Owl Creek Mountains, across the basin and out through a canyon gap in the Big-horn Mountains.

The Wyoming Basin has some interesting man-made development despite its semiarid climate Dams and reservoirs in the mountain-fed streams control considerable irrigation water The underlying basin layers have yielded much petroleum. The first transcontinental rail route, the Union Pacific, extends through the southern basin and is paralleled by transcontinental highway U S. 30 The Oregon Trail was notable among several earlier trails which passed through this Rocky Mountain intermontane basin.

Northern Rockies. The Northern Rockies of Idaho and western Montana in the United States and of British Columbia and southwestern Alberta in Canada are a varied and complex mountain land, but contain certain consistent characteristics. First, the mountain front swings northwestward and roughly parallels the Pacific Coast in this narrowing part of the Cordilleran region. Second, piedmont hogbacks are largely missing at the junction with the plains because most of the linear Rocky Mountains have been strongly pushed, or overthrust, from the west to override lower formations. Some of them override the sedimentary layers of the Great Plains Third, outlying small domes, eroded into hill and low mountain country, dot the plains of bordering Montana, but are infrequent or missing in Canadian plains. Fourth, the Canadian Rockies are a narrow belt (70–80 mi) of roughly parallel ranges separated by the Rocky Mountain Trench from the interior plateaus of Canada in northern British Columbia; but the mountain region broadens complexly westward of this trench in southern Canada.

Next to the Rocky Mountain Trench lies the Purcell Range and then the Selkirk Range, both cut off diagonally on the north by the trench. A mass of mountains sometimes designated the Columbia Range extends as a broad band of highlands some 300 miles southward into the United States at the Columbia River. Where the Fraser Plateau pinches off at the south, the Columbia mountain mass closely adjoins the Cascade Range of the Coast Mountains at the Canada–United States frontier. All of these mountains in both the United States and Canada are strongly marked by mountain and valley glaciation, and many of the valleys contain elongated finger lakes in their unevenly eroded glaciated valley bottoms The northern Selkirks contain majestic mountain scenery and a variety of glacial features that are set off in the Glacier National Park, just west of a larger famous ...

... system The Canadian National Railway passes through Yellowhead Pass, at 3700 ft, in Jasper National Park not far from Mount Robson, which has a 13,068-ft summit Farther south, the Canadian Pacific uses remarkable spiral tunnels to go through Kicking Horse Pass in Banff National Park. Motor highways now penetrate these same pass routes.

Diversity of mountain character and a few outstanding man-made features mark the three-part United States portion of the Northern Rockies. The Waterton Glacier International Peace Park ex-

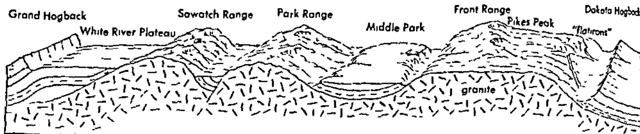


Fig. 13. *Structure-topography diagram of the Southern Rocky Mountains. (From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)*

highland relief. This article uses a classification with three or four levels of category to outline the highland characteristics. The largest parts may be designated as three systems: Rocky Mountain system, of ranges extending in various parts on the eastern margin of the Cordilleran region from New Mexico to Arctic shores of Alaska; intermontane plateau system, westward of the Rockies and extending even further (from southern Mexico to Alaska); and Pacific mountain system, from the Aleutians to peninsular Lower California. The outstanding physical characteristics of numerous distinctive parts are then briefly discussed under italicized headings, such as Northern Rockies, Columbia Plateau, and Sierra-Cascade-Coast mountains; but their notably different parts are then discussed in separate paragraphs, or they are numbered, such as (1) Sierra Nevada section or (2) Yukon Plateau section.

Rocky Mountain system. Four masses and ranges of mountains upstand above the interior and Arctic plains regions of North America from north central New Mexico to Alaskan Arctic shores, north of Bering Strait. The divisions have been designated Southern Rocky Mountains, Middle Rocky Mountains, Northern Rockies, and Arctic Rockies.

Southern Rocky Mountains. Although mountain structures and some rugged terrain stand between the North American plains and the Cordilleran plateau system to the south (as in the Guadalupe Range in the United States and the Sierra Madre Oriental of eastern Mexico), the Rocky Mountain system is considered to have its beginning in this area of massive rugged ranges with intervening valleys and occasionally more open basin parks Floors of the latter contain some remaining downwarped sedimentary remnants and are littered with recent alluvium. A great up-arched or wrinkled structure appears to have been uplifted one or more times and degraded by erosional processes. During such erosional processing a peneplain developed, truncating the wrinkled structures The peneplain was surmounted in places by upstanding remnants or monadnock ranges and peaks, left as reminders *of once-higher overlying* materials These stand out as conspicuous eminences in the Front Range and particularly in such as Longs Peak and Pikes Peak (see diagram sketch of Fig 13). On the eastern margins, the Rockies in Colorado are flanked by a series of valleys and parallel hogback ridges left by differential erosion in exposed edges of the younger sedimentary rock layers, where they are upturned against the Rocky Mountain front. This forms the Colorado Piedmont, and in localities of weird forms etched into red sandstone, the Garden of the Gods (near Colorado Springs). The Grand Hogback is the conspicuous piedmont feature of the western side.

The main trend of the ranges and valleys or parks (except for the more volcanic mass of the San Juan section on the southwest) is so rugged, unbroken, and so nearly north-south as to be more of a barrier than the Middle and Northern Rocky Mountains. Santa Fe, a town at the southern end of the region, has long been associated with through trails and routes between the Great Plains and the West A 12-mile Moffat tunnel pierces the Front Range near Denver, at the piedmont margin Somewhat southward a rail route uses the narrow Royal Gorge where the Arkansas River notches through to the Great Plains. The old Union Pacific route crosses the narrow northern part of the region to make a relatively low and short way through.

This mountain area contains a variety of physical features and resources, the development of *which is made difficult by isolation, irregular* and patchy distribution (as for forest, grass ranges, cultivable land), and by the fixed and exhaustible amount of such resources as minerals. Many a famous mining place has thus become a ghost town. A few places of multiple function and strategic location in the area continue as cities and towns: on the east, generally at places where plains routes funnel into canyons or passes through the highlands, are Cheyenne, Boulder, Denver, Colorado Springs, Pueblo, and Trinidad, but Santa Fe is situated at the south

Middle Rocky Mountains A large embayment of Great Plains structures indents the Rocky system at the north end of the Southern Rocky Mountains to form a basin floor, with scattered hills and mountains. This Wyoming Basin floor is actually a series of smaller structural basins between scattered domed, arched, or folded and faulted structures which are eroded into a variety of hilly and mountain patterns. The northwest-southeast Wind River Range, a linear folded and faulted anticlinal mass, extends from the Yellowstone section well into the center of the Basin. The encircling mountain terrain has structures and develo landforms

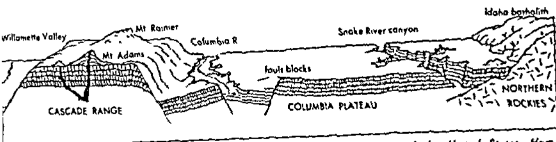


Fig 14 Structure-topography diagram of Columbia Plateau and the Cascade Range (From A. K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

1 Cook Inlet-Susitna and Copper basins. The Cook Inlet-Susitna Basin with Anchorage as its urban center, supports a growing population, and petroleum is found in pools entrapped in that part of the basin associated with Kenai Peninsula A highway from Anchorage uses the Matanuska Valley route between the Talkeetna and Chugach mountain areas to pass eastward into the elevated and near-treeless Copper Basin. Here it joins the Richardson Highway, on its route north from the Pacific coastal port of Valdez toward Fairbanks; and the Tok Passway road swings northeastward toward the Alaska Highway near the Canadian border.

2. Pacific coast ranges Four units—Kodiak Kenai, Chugach, and St Elias Ranges—lie on the margins of southern Alaska and, with the exception of the Panhandle and Kodiak and Afognak Islands, are roughly mountainous with maximum elevations of 4000 ft and irregular shorelines. The Kenai and Chugach Ranges are compactly rugged and much glaciated mountains of the south Alaskan coastal zone The many mountain glaciers are fed by the heavy precipitation from the warm and moist Pacific air impinging on these shores. Warm Pacific waters also induce mild winters and ice-free harbors along the extremely irregular shores, and fine evergreen coastal forests stand on many lower slopes and coastal margins Between the Chugach mountains and the Alaska Range and to the east of the Copper Basin is a group of mountains much h...

The magnificent mass of ice and mountains in the Saint Elias Range is notably surmounted by groups of snow-clad peaks some of the highest in Canada and the continent. (For a discussion and illustration of these mountain and piedmont glaciers, see GLACIER) To the south the character of the Pacific coast ranges is markedly different.

3 Inland passage section. A strongly glaciated coastal moun- ... lands ar ... Alaska ... of tidew... the reappearing Pacific Troughs here take the form of a sheltered waterway in this inland passage section The sheltering mountainous islands are, from north to south, the Alexander Archipelago, Queen Charlotte Islands, and the Vancouver Range.

Canadian coast mountains. Unlike the offshore Pacific coastal ranges, these are the ice-eroded and intricately fiorded margins of great igneous rock masses at the western margin of the mainland Warm Pacific waters and moist air masses bring maritime moderation and beautiful evergreen forests to many of the lower slopes and shores, but upland winters are cold and highland snows are heavy A few upland ice fields remain inland from the Alaskan panhandle In many places, the upland surface between the U shaped glaciated valleys and on the interfiord ridges appear to join rather evenly with a great backland plateau which was recently and widely scoured by ice cover. This elevated surface is occasionally surmounted by peaks and seems to be somewhat continuous with the upland surface in many of the western parts of the Canadian interior plateaus.

Intermontane plateau system. In Canada these interior plateaus present, on the west, much of fairly even upland surface occasionally surmounted by peaks and scattered ranges and here and there deeply gashed by glaciated valleys, many with long lakes. Some of the dissecting valleys are straight in seeming response to structural weaknesses but many are dendritically irregular in pattern Relief and elevation become rougher and higher on the eastern margins to become commonly mountainous, but less linear in pattern than in the adjoining Rocky Mountains This condition has been mentioned before in the cases of the Ogilvie and Selwyn mountains, adjoining the Mackenzie Mountains on the eastern margin of the great Yukon Plateau. This Yukon Plateau reaches more than 400 miles from southeastern Alaska into Canada The remainder of the 1000 miles, north-south, of these plateaus bears names of the principal draining rivers The Stikine section is in the middle and the Frazer Plateau in all the southern portion of these elevated, rugged, and partly mountainous Cordilleran uplands.

South of the Canadian border, in the United States and Mexico, are four major units of the intermontane plateau system: Columbia Plateau, Basin and Range province, Colorado Plateaus, and Mexican Highland

Columbia Plateau Huge outpourings of lava exuded at various times have built up thick layers to form at some 3000 ft a sea of lava, overtopping all but a few of the previously existing features of inland basin topography. This is observable along

tends into the Rockies of Montana and southern Canada. Igneous intrusions have brought mineral resources to parts of these regions, and such outstanding developments as those in the Butte, Anaconda, and Helena districts result. Three transcontinental rail routes, the Great Northern, Northern Pacific, and Milwaukee, swing through the Montana part of the Rocky Mountains, but even the two more southerly routes are diverted far enough northward between Missoula and Spokane to avoid the nearly impassable terrain of the Bitterroot section Forest and mineral resources characterize both the Bitterroot and Salmon River Mountains, but isolation and rugged terrain leave them largely undeveloped or in national forest reserves for scenery and possible future use.

Arctic Rockies. Predominantly parallel linear ridges and ranges of Rocky Mountain type make a great curve between the Mackenzie Lowland of Canada and the elevated and rugged interior plateaus Because of greater proportion of igneous intrusive masses in the plateau structures their highland and mountain terrain is less regularly linear. This is a point illustrated by the difference between the curving Mackenzie Mountains and their easterly parallel outlier, the Franklin Mountains, and the less regularly arranged mountain masses of the Ogilvie and Selwyn Mountain sections of the interior plateaus of Canada Similar but less strongly developed folds and domes appear in the plains just east of the Franklin Mountains and furnish the basin structures entrapping the petroleum resources developed during World War II at Fort Norman Wells. The oil was refined and sent by a great pipeline and parallel service road developed through the mountains and plateaus to Pacific tidewater at Skagway in the northern panhandle of Alaska. The line was later abandoned and the metal pipe reclaimed for reuse elsewhere.

The Alaska Highway finds a way through near the Liard River Gap to pass from Fort Nelson into the Liard Plain at the north end of the Rocky Mountain Trench. Few other easy ways through the Arctic Rockies are known until the low gaps at the ends of the Richardson Mountains but these are undeveloped. At the southeast end is a broad gapway of the Bonnet Plume Basin where the Peel River flows through to join the lower Mackenzie. There are low saddles at the northwest end of the Richardson Mountains connecting the coastal zone of Mackenzie Bay with the watery Porcupine Plains and thence to the Yukon Valley.

The Brooks Range is high and compactly rugged mountain terrain of rather complex Rocky Mountain type of structure through the first three-fourths of its westward extent to Kotzebue Sound and the Arctic coastal zone. Here the Brooks Range is a barrier to easy passage and a barren zone of transition between central Alaska with some boreal forest and the treeless tundra land of the Arctic coastal plain The western quarter becomes hill and low mountain country interspersed with low passes and a few upland basins.

Central Alaska. Plains interspersed with several areas of hill and low mountain country characterize all save one part of Alaska between the Brooks Range to the north and the great Alaska-Aleutian Ranges to the south. Plains are nearly flat in areas of recent glacial outwash and in the alluvial bottoms of some of the larger rivers, such as the Yukon, Kuskokwim, and Tanana; and in the great deltas of the Yukon and Kuskokwim Rivers they are very flat.

In areas where marine influence contributes to chilly cool summers the vegetation is tundra, as in the west and northwestern parts. Northern coniferous and mixed forest, however, survives in inland parts having a more continental short warm summer except in higher parts, such as in the compact protrusion of the interior plateaus between the Yukon and Tanana Rivers in southeastern central Alaska.

Pacific mountain system, Alaska, Canada. Save for the one protrusion mentioned above, the interior plateaus are in Canada, and will be considered in this connection in a later section The Pacific mountain system here has two main parts, the Sierra-Coast mountains, and Pacific coast ranges, but these are separated in several places by areas of basin and trough, designated Pacific Troughs

Alaskan Sierra-coast mountains. The arcuate chain of the Aleutian Islands appears to be only the upper portions, exposed above the level of the sea, of a great volcanic mountain range. Despite the lack of much level land, dangers from volcanic eruption and earthquake, and the stormy chill climate throughout the year, ways are found to develop usable landing fields and aviation way stations along the great circle routes, and surface ship facilities continue in natural harbors at such places as Dutch Harbor and the non-Aleutian but nearby Kodiak Island

Similar conditions of rugged volcanic mountains, marked with many active and quiescent volcanoes, caldera, and zones of gaseous vents, such as the famed Valley of Ten Thousand Smokes, stand in a curve with ragged shore on the southern side of the Alaska Peninsula. The treeless plain on the northern side broadens along Bristol Bay, of the Bering Sea, to the break in terrain near the large Lake Iliamna

Great compact masses of high rugged mountains, occasionally punctuated with volcanoes, reverse the curve of the Aleutian arc convexly northward around all the south Alaskan mountain and basin country Broadest northwest of Anchorage and the Cook Inlet, the range culminates in height somewhat northward at the spectacular Mount McKinley (20,200 ft) The Alaska Railway uses an elevated pass, just east of Mount McKinley National Park, in going northward from ice-free ports of Seward and Whittier via Anchorage to Fairbanks in central Alaska. The range narrows and has a few lower passes to the east through which two roadways pass from Copper Basin northward to the Alaska Highway in the Tanana Valley.

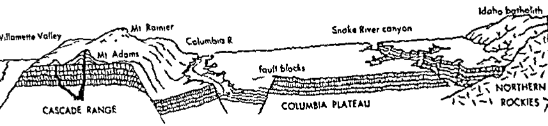


Fig 14 Structure-topography diagram of Columbia Plateau and the Cascade Range (From A K Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

1. Cook Inlet-Susitna and Copper basins The Cook Inlet-Susitna Basin, with Anchorage as its urban center, supports a growing population, and petroleum is found in pools entrapped in that part of the basin associated with Kenai Peninsula. A highway from Anchorage uses the Matanuska Valley route between the Talkeetna and Chugach mountain areas to pass eastward into the elevated and

Tok Passway road swings northeastward toward the Alaska Highway near the Canadian border

2 Pacific coast ranges Four units—Kodiak, Kenai, Chugach, and St Elias Ranges—lie on the margins of southern

Chugach Ranges are compactly rugged and much glaciated mountains of the south Alaskan coastal zone The many mountain glaciers are fed by the heavy precipitation from the warm and moist Pacific air impinging on these shores Warm Pacific waters also induce mild winters and ice-free harbors along the extremely irregular shores and fine evergreen coastal forests stand on many lower slopes and coastal margins Between the Chugach mountains and the Alaska Range and to the east of the Copper Basin is a group of mountains much

The magnificent mass of ice and mountains in the Saint Elias Range is notably surmounted by groups of snow-clad peaks some of the highest in Canada and the continent (For a discussion and illustration of these mountain and piedmont glaciers, see GLACIER) To the south the character of the Pacific

coa

lan

Al

indented with arms of tidewater. The reappearing Pacific Troughs here take the form of a sheltered waterway in this inland passage section The sheltering mountainous islands are, from north to south, the Alexander Archipelago, Queen Charlotte Islands, and the Vancouver Range.

Canadian coast mountains. Unlike the offshore Pacific coastal ranges, these are the ice-eroded and intricately fiorded margins of great igneous rock masses at the western margin of the mainland. Warm Pacific waters and moist air masses bring maritime moderation and beautiful evergreen forests to many of the lower slopes and shores, but upland winters are cold and highland snows are heavy. A few upland ice fields remain inland from the Alaskan panhandle. In many places, the upland surface between the U-shaped glaciated valleys and on the interfiord ridges appear to join rather evenly with a great backland plateau which was recently and widely scoured by ice cover. This elevated surface is occasionally surmounted by peaks and seems to be somewhat continuous with the upland surface in many of the western parts of the Canadian interior plateaus.

Intermontane plateau system. In Canada these interior plateaus present, on the west, much of fairly even upland surface occasionally surmounted by peaks and scattered ranges and here and there deeply gashed by glaciated valleys, many with long lakes. Some of the dissecting valleys are straight in seeming response to structural weaknesses but many are dendritically irregular in pattern. Relief and elevation become rougher and higher on the eastern margins to become commonly mountainous, but less linear in pattern than in the adjoining Rocky Mountains. This condition has been mentioned before

mountains,

the eastern

This Yukon Plateau reaches more than 400 miles from southeastern Alaska into Canada. The remainder of the 1000 miles, north-south, of these plateaus bears names of the principal draining rivers. The Stikine section is in the middle and the Frazer Plateau in all the southern portion of these elevated, rugged, and partly mountainous Cordilleran uplands.

South of the Canadian border, in the United States and Mexico, are four major units of the intermontane plateau system. Columbia Plateau, Basin and Range province, Colorado Plateaus, and Mexican Highland.

Columbia Plateau Huge outpourings of lava exuded at various times have built up thick layers to form at some 3000 ft a sea of lava, overtopping all but a few of the previously existing features of inland basin topography. This is observable along

the few canyon sides where main streams, the Columbia and Snake Rivers, have cut into and through the overlying lavas to reveal underlying structures. In a few places faulting causes scarps and some tilted block mountains. Various mountains stand as an "island" area surmounting the upland surface in the Blue Mountains section. Volcanic craters, called Craters of the Moon, and youthful lava surface mark the southeastern arm of the Snake River plain in southern Idaho. An elevated basin part on the south contains volcanic features, some block ranges, and basins of interior drainage in the Harney section of the southwest.

Deeper soils, rolling surface, and some areas of irrigation are reflected in agricultural development, mostly in the northern parts. Here are the great machine-farmed wheat lands of the Walla Walla Plateau of Washington and Oregon. Irrigation and fruit raising are famous in the Yakima Valley district and in parts of the broad bottom of the deep Columbia River valley, which is also much used as a transportation route.

Basin and range province. Five parts characterize this largest (over 300,000 mi²) division of the intermontane plateau system of Cordilleran North America. Mountain and basin, or desert bolson, topography and the prevailing arid and semiarid climate are the widespread unifying attributes of this area.

On the north is the well-known Great Basin. Here the repeated pattern is of block mountain ranges, mostly oriented north and south and in varying stages of erosional destruction These stand between bolson basins of interior drainage that are filling with the erosional offscourings from the adjoining highland ranges (For a consideration of the details of landform features, *see* DESERT EROSION FEATURES.) The Great Salt Lake is the largest of the desert playa lakes. Nearby Salt Lake City and its oasis agriculture are outstanding achievements based on the use of waters of the mountain streams of the adjoining Wasatch Range.

To the southeast in Arizona and New Mexico are open-basin sections of these dry-land regions. In contrast to the closed basins of interior drainage, these basins drain to the Colorado or Rio Grande rivers, or to the sea. The basins are occasionally marked by oasis settlements and a few urban centers, such as Tucson This large division is margined on the east by the Sacramento section where north-south block-faulted ranges and elongated intervening basins run south from the Rocky Mountains to border the Great Plains province Several of the past trails and the present more southerly rail and highway routes pass through these dryland parts of New Mexico and Arizona on their way to the West Coast.

Some of the deeply filled basins and the maturely dissected mountain ranges are now of considerable interest, for example, the Tularosa Basin west of the high (1200+ ft) Sacramento Range. This basin contains the White Sands National Monument (great expanses of white dunes and gypsum) and the White Sands proving grounds for development of atomic devices Several outstanding highland or mountain dam and reservoir facilities, such as Roosevelt Dam on the Salt River, Parker Dam and Boulder Dam (Lake Mead) on the Colorado River, and Coolidge Dam, furnish scenic and recreational interest and valuable waters for irrigation and for a few urban developments in the valleys and basin floors.

The Sonora Desert and the Salton Trough, of California, Arizona, and northwestern Mexico, extend the Basin and Range province some 1700 mi to the south. On the north and northwest, this desert region is separated from the dryland coast by numerous ranges and contains repeated range and basin desert, including the depressed basin of Death Valley east of the Sierra Nevada between the Mojave Desert of California and the basin and range country to the north. The Salton Trough is a northern extension of the linear depression of the Gulf of California; its area north of the Colorado River "delta" is below sea level, making it a troublesome place of variable flooding and desiccation Winds once transported deltaic sands into the dune area of Gran Desierto in Mexico. Today controlled irrigation waters of the Colorado are released to support the large Imperial Valley farming area of California and a lesser but growing development in Mexico The rest of the Sonoran area is a narrowing dryland transition between the Mexican Highlands and the Gulf of California on the west. About 100 mi wide in the north, it resembles the basin and range country of the deserts in the United States. Southward it is progressively narrower and a little less arid A large part, however, is an arid region of nearly buried hills and low mountains protruding above the alluvial litter washed from the Mexican Highland during desert downpours and mountain rains.

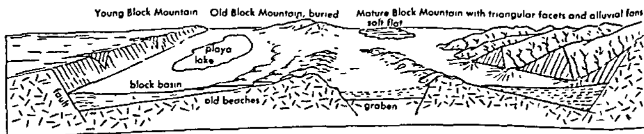


Fig. 15. Structure-topography diagram, east-west through the Basin and Range province. (From A. K. Lobeck, *Physiographic Diagram of the United States*, Hammond, 1957)

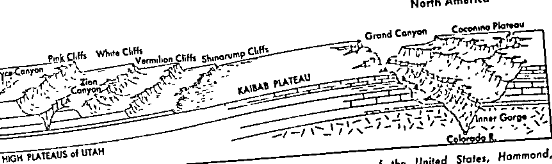


Fig 16. Structure-topography diagram of a part of the Colorado plateaus (From A K Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

Colorado plateaus Great areas of relatively undisturbed and almost horizontal beds of rock lie thickly upon an underlying basement complex between the Southern Rockies and the Great Basin. Great flattened summit areas appear in these elevated uplands, broken here and there by structural ... particularly ... canyon val... exposes the ... below which ... ape into the ... Canyon

... other parts A few faults result in conspicuous scarps Scattered sharp domes are in varying stages of erosional destruction into hill and mountain areas From the margins of these, and from washing and erosion of up- or downwarped plateau sections, hogback ridges are developed, and dryland escarpments and table or mesa remnants are conspicuous. Scattered volcanoes and lava outpourings punctuate a few places. Although mostly semiarid, the higher surfaces support fine forests which are little used in this great region away from the settled and developed parts of the country

Mexican Highland Essentially five coextensive areas and one separated part make up the Mexican portion of the North American Cordillera The long and complexly faulted peninsula of Lower or Baja California is the southern extension and end of the Pacific mountain system The mainland highlands have a highland mass of four parts with predominant north south trend and other characteristics like those of western United States and Canada These are (1) the elevated and dissected lava (rhyolite) plateau cut into rugged mountain terrain on the west to form the Sierra Madre Occidental, (2) folded mountains in the Sierra Madre Oriental standing up southward from the complex Big Bend area on the Rio Grande, between the Gulf Coastal Plain and the two interior parts of the upland, (3) a more northerly area of block ranges and basins similar to those of the United States, (4) southward of low ranges (curving west from the Oriental to the Occidental Sierra Madre), a great elevated basin floor (although locally called Central Mesa) marked here and there by hilly areas resulting from dissection of old volcanoes and characterized by considerable settlement and scattered cities, and (5) a neovolcanic plateau, abruptly terminating the North American Cordillera on the south. This plateau extends from the Pacific, near Tepic, to steeply sloping margins close to the shores of the Gulf of Mexico near Veracruz, and the Tuxtlas volcanoes stand conspicuously at the shore of the coastal plain just south of Veracruz This great upland plateau is considered by some authorities to be on an east-west shear zone in the underlying rocks, with the great number of high, youthful volcanoes arranged generally in correspondence with the zone of structural weakness and lateral displacement The elevated upland has a moderate climate and is somewhat more humid than the dryland plateaus to the north, and it has become the heart of the nation and the site of Mexico City, the capital. Beyond lie the ranges and gapways of Central America

Pacific mountain system, southern. A previous section outlined the character of the Pacific mountain system in Alaska and western Canada; this ... and basin character in ...

... tains, the Pacific troughs, and the Pacific coast ranges, extending in that order between the Cordilleran intermontane plateau system and the Pacific Ocean.

Sierra Cascade mountains Two great tilted block mountain sections, that of Lower California and the Sierra Nevada of California, and that of the Cascade section of Oregon and Washington, comprise this nearly unbroken unit. The rough, complexly faulted section of Lower California bears the same relationship to the basins and ranges of the Sonoran Desert area as the Sierra Nevada does to the basins and ranges of the Great Basin, except that the adjoining Mexican basin contains the Gulf of California (see Fig 17). Lower California is largely a desert mountain region, little inhabited and scarcely developed, even for its known minerals. California's Sierra Nevada presents considerable contrast

Although its past history is complicated, the Sierra Nevada block range is understandable in relatively simple aspects of its structure, form, and related physical features. A great peneplaned mass was faulted and the block tilted with the old surface on the gentler slope to the westward, whereas

the abrupt eastern slope presents some 3000 ft of barren gullied scarp above the adjoining Great Basin The crest is high and ruggedly etched by mountain glaciers The gentler western slope is more humid and contains great evergreen forest belts, at elevations between 6000 and 9000 ft. Although an isolated timber resource, the forests are great regulators of runoff into streams for dryland irrigation in the Central Valley of California. The peneplaned rocks contain mineral veins, subject to erosion by the mountain streams. Thus were deposited the gold placer deposits (of stream sands and gravel) which initiated the Gold Rush of 1849. Later the gold and other minerals were mined from their lodes and veins. The great block range extends some 400 mi to end near the northern border of the state of California.

The Cascade Mountains begin as a spectacular array of volcanoes and volcanic materials north of the Sierra Nevada and gradually change to a massive horstlike uplift with scattered volcanic peaks in their more northern parts (see Fig 14). These mountains are a formidable barrier through most of their extent to the Fraser River valley of southern Canada Few through routes cross except via the valley of the Columbia River. Mild marine air masses of the Pacific coastal zone bring rainy climate and some of the finest fir forests of the continent to the western slopes, but the eastward slopes become drier, and the rain shadow means dry climate on the Columbia Plateau, although mountain streams bring water for such irrigation development as that of the Yakima Valley in Washington.

Lowland troughs and coast ranges In spite of some complex differences and numerous local variations, the troughs and coast ranges from southwestern Canada to Mexico present a generally bold and simple cordilleran coastal-zone pattern. A trough extends north-south in the Georgia depression between the Vancouver Range and the Canadian Coast Mountains to connect tidewater via Georgia Strait between the Inland Passage and the waters of Puget Sound This Georgia depression is commonly considered to be a part of the Puget Sound section.

The coastal zone of Washington and Oregon is composed of (1) the Puget Sound section of lowlands running southward two-thirds of the distance to the California border, and (2) coast ranges in the Olympic Mountains, the Oregon Coast Range, and the broader Klamath Mountains to the south of the lowland sections. The Juan de Fuca Strait runs in from the Pacific to Puget Sound and thus separates the Vancouver mountains from the deeply dissected and compactly rugged mountain mass of the Olympic Mountains. With general ridge summits at 5000 ft and a few surmounting peaks to about 8000 ft, this upland induces heavy precipitation from the mild marine air masses. Heavy evergreen rainforest grows to great heights (many trees more than 200 ft) from a mossy forest floor of tangled undergrowth. Much lumbered on the margins, a considerable portion of this region is now being preserved as a national park. The mountains are ringed by lower land, but there is a hilly plain gap in the coast ranges, through which flows the Chehalis River. To the south are hills and low mountains cut by the lower Columbia Valley.

Low mountain country marks the Oregon Coast Range. With mild marine climate, these are some of the wettest (140–150 in./yr) and most heavily timbered parts of the continent. Southward of about 44°N the terrain grows rougher and the mountains broaden, called here the Klamath Mountains, which reach from the coast inland to the volcanic Cascades.

The north-south Puget Sound section presents two distinct parts southward of the Georgia depression and the landward end of Juan de Fuca Strait. First is the Puget Sound lowland plain containing many arms of the Sound Agriculturally important, this heavily populated region is also becoming important in industry and commerce in such cities as Seattle, Everett, Tacoma, and Olympia. Southward the basin floor grows hilly to form a divide between this and the next main lowland parts.

The Cowlitz valley area lies north of the Columbia River, but the greater part of this southern lowland lies in the Willamette Valley region of Oregon. This is predominantly agricultural land with some of the greatest population of the western United States. Portland, however, with its tidewater connection on the Columbia and its command of overland routes, is growing to rival Seattle in urban development.

Only the broad physical plan of the cordilleran coastal zone in California may be outlined in this account. On the north the Klamath Mountains continue compactly rugged, well forested, and a considerable barrier to north-south travel. The only through route of rail and highway to the Willamette Valley winds through via Grants Pass. Southward the Klamath mass of mountains gives way on the east to the volcanic transition between the Sierra Nevada and the Cascades In this vicinity Shasta Dam and reservoir regulate the waters of the lowland Sacramento River Southwest of the Klamath area, the Coast Ranges commence their low-mountain pattern of parallel ridges and valleys between the Pacific shore and the length of the Great Valley Many of the valleys and ridges come to the coast at a slight northwest angle, so that valley shores are low and commonly somewhat embayed, whereas the mountain coastal ends are repeatedly promontoried Some of the valleys are agriculturally important, especially in vines and fruit just north and south of the San Francisco area Wooded in the northern parts, the Coast Ranges become covered with brushy chaparral and grass, or are progressively more barren, in reflection of the mediterranean dry-summer climate and the semiarid lands to the southward. Frequent fogs on this coast give way to sunnier lands a few miles inland, particularly in the Great Valley.

The alluvium-filled trough depression of California is 400 mi long and [illegible] 50 mi w[illegible] [illegible] areas under m[illegible]

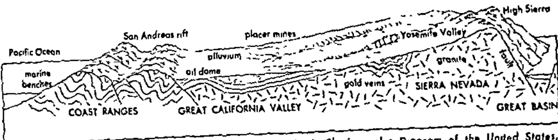


Fig 17 An east-west structure-topography diagram through the Sierra Nevada, Central Valley depression, and Coast Ranges of California (From A K. Lobeck, Physiographic Diagram of the United States, Hammond, 1957)

south) semiarid climate are used for agriculture. Some grazing gives way to extensive machine cultivation of grains, then to truck farming and intensive fruit and vine growing Valuable waters are obtained for irrigation, especially from Sierra streams and through the widespread tapping of artesian waters in the alluvial fill of the basin. The Great Valley produces oil and gas in scattered places, and the whole area is dotted with small urban communities. The greatest proportion of the oil, however, is obtained from the folded Coast Range structures to the southward and even beneath the coastal sea waters

The great urban centers of central California are localized where the rivers draining the Great Valley join and flow west through the Coast Ranges to the Pacific This valley area appears recently depressed and the lower parts are drowned by the waters of the Golden Gate, San Francisco Bay, and, to the northeast, San Pablo Bay The great urban complex includes San Francisco, on the south of the Golden Gate and the Oakland and Berkeley communities near the eastern end of bridges across San Francisco Bay

The Great Valley ends on the south in lower ranges curving southwestward from the end of the Sierra Nevada to the Coast Ranges where they adjoin the Los Angeles Ranges

The Los Angeles Ranges are a somewhat complicated system of ranges and masses of mountain terrain The general trend is eastward and inland from the coast, then curving southward around the end of the tilted block structure of Lower California Some mountain structures of the continental shelf appear in several islands off these Pacific shores. On the mainland are numerous interridge valleys and a few lowlands In the northern part of the largest of the lowlands, where it opens to the Pacific, is Los Angeles This metropolitan area extends inland and also northward, where it reaches into highland and valley areas The surrounding lowlands and valleys are cultivated in citrus and other fruit The semiarid climate and the huge demand for water in the cities require transport of water from such distances as Owens Lake and the Colorado River by means of long overland aqueducts

Old alluvial slopes at the western base of the faulted mountains, sometimes termed composite deltas, form a lowland shore to the south of the Los Angeles area Here on a bay and harbor behind a coastal bar lies the city of San Diego The climate being arid, there is complete dependence on irrigation for cultivation of fertile soils. Water supply continues to be of major importance, although San Diego now also receives water from the Colorado River. South of California only a semblance of the Coast Ranges appears along the western margins of Lower California, or else they are totally submerged by the Pacific.

CENTRAL AMERICA

This narrowest part of the North American mainland has predominantly mountainous landforms, but it is geologically and structurally different from the great ranges, basins, and plateaus of cordilleran western North America It is perhaps best considered as a composite of at least five rather separate and different parts, more closely related to the Antillean mountain system.

Sierra del Sur. Despite the more east-west trend of complex structure and lessened predominance of volcanoes, the ruggedly mountainous land of Sierra del Sur gives much similarity of appearance [illegible]

[illegible] levels of the rugged terrain *reflect* rather complexly the vertical differentiations and horizontal zoning in vegetation and agriculture that mark highlands of the [illegible]
and [illegible]
mak[illegible]
shor[illegible] *roads*, some rail lines, and airways penetrate the settled basins or pass *over* and through the rough highlands to connect the centers of the *volcanic* plateau of Mexico with the Pacific seashore resorts (as near Acapulco) or with the lowlands of the Gulf Coastal Plain *or* the low passway through the isthmus of Tehuantepec. The southeast of the *highland* terminates in an abrupt and rough fault scarp dropping to the lowland of Tehuantepec.

Tehuantepec-Honduras section. This *is* an *area of rough* mountain terrain with parallel ranges trending, in a flattened S curve, generally *from west* to east Much of this on the north is semiarid back country in the narrow valleys and margins. *It is* therefore much less settled and developed than the basins in the volcanic region to the southward. Rough basin country, 1500–3000 ft elevation, in the middle and upper course of the Grijala River sys-

tem supports, however, a considerable density of rural population. The lowland passway appears as a hilly-floored graben, downdropped between these two highland masses in such a way as to connect, at less than 800 ft elevation, the Gulf Coastal Plain with the narrow but well-developed Pacific coastal plain to the southeastward. Although it is longer than most others, this pass route presents some possibility as an alternative or additional route to that of the Panama Canal. A railway now traverses the gap from Gulf to Pacific.

Volcanic upland. Volcanoes and volcanic ash and some lavas characteristically mark the backbone highlands of southern Guatemala, Salvador, Honduras, and Nicaragua. These highlands and their volcanic peaks (in places in a semblance of one, two, or more rows) rise abruptly back from the coast or narrow coastal plains on the Pacific side which is also the drier tropical exposure. The upland slopes gradually downward toward the other side in a series of long eastward ridges separated by widening, alluvial-floored valleys. These merge with the low, rainy, tropical plain of Mosquito Coast along the Caribbean, but the submerging ribs of the ridges appear as en échelon irregularities of the coast or stand as scattered offshore islands

A seemingly continuous depression definitely separates the main volcanic upland in Nicaragua from a similar but elongated highland along the Pacific coastal zone. Actually this extends more or less unbroken from the plain drained by the Rio San Juan from Lake Nicaragua to the Caribbean, to the Gulf of Fonseca in the Honduran coastal zone. This same structural depression seems also to reappear in the west-east depression drained by the upper Rio Lempa in Salvador.

A potential transisthmian waterway route might utilize the Rio San Juan to the large volcano-dotted Lake Nicaragua. Several places appear potentially feasible locations for canal connection to Pacific waters

Coastal lowlands. Alluvial offscourings form long, comparatively narrow, and flat-lying plains on both sides of these Central American highlands. That of the western side is composed of a large proportion of volcanic material from the volcanoes of the highland and is somewhat drier in the high-sun period. Both are tropically warm to hot but the eastern plain area is rainy throughout the year The rainy eastern coastal zone has been much developed in banana plantations, but these have declined in recent years particularly from the damaging ravages of plant disease and blight. Many plantations have been established recently on the drier western plains and some are now connected by rail for Caribbean shipment on the eastern coast.

Panamanian arc. On the Pacific side of the Nicaraguan *lowland gapway*, a narrow volcanic upland surmounted by numerous volcanic peaks swings southeastward to join the variably rough mountain terrain that dominates the feasibility of various settlements and transisthmian travel despite scattered areas of lower land. Highland settlement in the volcanic mountains and basins of Costa Rica has been progressive and expanding in recent times, but the wetter and warmer lowlands continue to present more difficulties for development on the western side and remain largely unsettled on the Caribbean side. Panama's development is similar but lacks the stimulating highland and basin zones.

The Panama Canal passes through one of several nearby ways that were used for past land travel in the isthmus of Panama between Caribbean and Pacific waters. It uses a reservoir and lock system, in part to control local floods, in making the low (85-ft) canal way some 40 miles through the Panama Canal Zone, developed largely by the United States. Long an important place in travel and settlement plans since the coming of European men to the New World, the isthmus of Panama continues to be a place of some troublesome problems and widespread significance in the present day.

[C.V.C., E.C.P.]

Bibliography: W. W. Atwood, *The Physiographic Provinces of North America*, 1940; N M. Fenneman, *Physiography of Eastern United States*, 1938; N. M. Fenneman, *Physiography of Western United States*, 1931; A. K. Lobeck, *Physiographic Diagram of North America*, 1950, E. Raisz, *Landform Map of Alaska*, 1948; E. Raisz, *Landform Map of Canada*, 1949; E. Raisz, *Landforms of Central America*, 1957, E. Raisz, *Landforms of Mexico*, 1959

North Pole

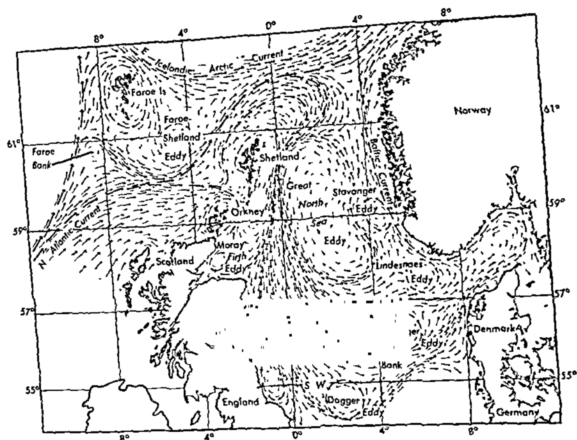
That end of the earth's axis which points toward the North Star, Polaris (Alpha Ursae Minoris). It is the geographical pole where all meridians converge, and should not be confused with the north magnetic pole, which is in the Canadian Archipelago The North Pole's location falls near the center of the Arctic Sea.

Being at the end of the earth's axis, which is inclined 23½ (23°45) from a line perpendicular to the plane of the ecliptic, the North Pole has phenomena unlike any other place, except the South Pole. For 6 months the sun does not appear above the horizon, and for 6 months it does not go below the horizon During this latter period. March 21–September 23, the sun makes a very gradual spiral around the horizon, gaining altitude until June 21, then it starts to lose altitude until it disappears below the horizon after September 23 The sun's highest altitude will be 23½°. As there is a long period (about 7 weeks) of continuous twilight before March 21 and after September 23, the period of light is considerably longer than the period of darkness

There is no natural way to determine local sun time because there is no noon position of the sun, and all shadows, at all times, fall to the south, the only direction possible from the North Pole. *See* GEOGRAPHY, MATHEMATICAL. [V.H F.]

North Sea

The North Sea overlies the European continental shelf between latitudes 51°N and 61°N. Its waters circulate freely with those [illegible]



Surface drift-currents in northern North Sea.

tic Ocean between Scotland and Norway and through the Straits of Dover The North Sea is a prolific fishery region Numerous fishing grounds have distinctive names, such as Dogger Bank and Fladen Ground A wide range of clupeoids, gadoids flatfishes, and crustaceans are commercially fished

The southern half of the North Sea floor is a plateau, mostly less than 40 meters (m) deep The northern half is a basin which deepens northward to the edge of the continental shelf at a depth of about 200 m There is a narrow submarine valley along the Norwegian coast with depths ranging from about 240 to 350 m

Atlantic oceanic water and continental (mainly Baltic) waters constantly flow into the North Sea The Atlantic Ocean water is warmer and saltier than that entering the North Sea from the Baltic In autumn and winter these waters mix to produce a characteristic water mass with intermediate conservative properties During other seasons a halothermocline exists at 30–40 m depth Occasionally, waters with Mediterranean or Arctic water mass characteristics invade the region Each of these water mass types has a characteristic fauna, chiefly plankton forms by which it may be recognized See HALOCLINE, THERMOCLINE.

The inflow of oceanic water through the Straits of Dover is relatively small; a greater volume enters from the north and becomes part of the prevailing circulation The surface current system in the northern art of the North Sea is shown in the accompanying illustration Subsurface and bottom currents set in similar directions except along the Norwegian coast. There a deep oceanic current moves south beneath a north-setting surface current Dynamically, and in their physicochemical characteristics, these several water masses vary seasonally, fluctuate annually, and undergo other changes over longer periods of time. Catastrophic variations, or those which cause unusual mortality among fish, are known to occur.

Mean minimum surface temperatures in February range from about 7°C in the northwest part of the North Sea to less than 2°C in the southeast part Mean maximum surface temperatures in August range from 11.5° to over 17°C from northwest to southeast, respectively. Extreme conditions of salinity do not coincide exactly with extreme conditions of temperature Salinity values for February (expressed in parts per thousand) range from nearly 35 25‰ to less than 32.00‰ and for August from more than 35.25‰ to less than 31 00‰ Other less conservative properties, such as phosphate, nitrate, and oxygen contents, vary seasonally and according to biological activity.

Marked tidal forces in the North Sea in conjunction with atmospheric disturbances produce surges which cause damage on neighboring coasts, particularly in the United Kingdom and the Low Countries See STORM SURGE. [J B.T.]

Bibliography: K. F. Bowden, Storm surges in the North Sea, *Weather*, 8.82–84, 1953; J H. Fraser, The plankton of the waters approaching

the British Isles in 1953, *Marine Research Service, Scottish Home Dept.*, (1):1–12, 1955; J. R. Lumby and G. T. Atkinson, On the unusual mortality amongst fish during March and April 1929, in the North Sea, *J. conseil, Conseil permanent intern. exploration mer.*, 4(3):309–332, 1929; J. B Tait, Hydrography of the Faroe-Shetland Channel, 1927–1952, *Marine Research Service, Scottish Home Dept.*, 2, 1957; J. B. Tait, The surface water drift in the northern and middle areas of the North Sea and in the Faroe-Shetland Channel, Fishery Bd. Scotland, *Sci. Investig.*, 1, 1937.

Northwest Passage

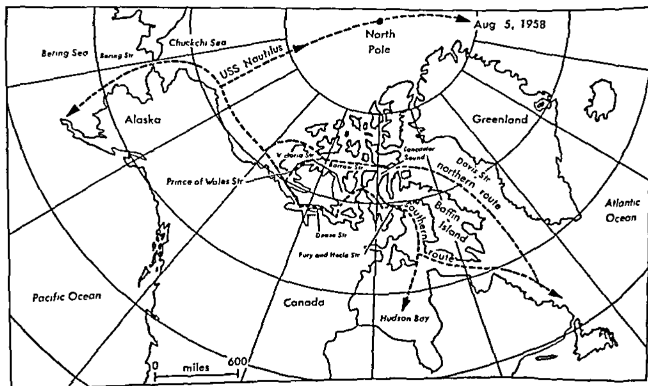
The northern sea route between the Atlantic and Pacific Oceans through the Canadian Archipelago. The entire route is frozen over in winter. As the ice cap retreats northward in summer two routes are opened (see illustration). The eastern approach through Davis and Hudson Straits is navigable in June, at which time the partially melted ice floes drift into the Atlantic Ocean on the Labrador Current. This approach remains open until November.

In the Canadian Archipelago the ice is landlocked (*see* ARCTIC AND SUBARCTIC ISLANDS). These passages remain closed until the ice melts. The tortuous southern route, close along the Canadian mainland coast, is usually passable from mid-August to mid October. In this period it has been traversed by a number of ships. The more direct northern route, through Lancaster Sound, the Barrow and McClure Straits, has never been found entirely ice-free. This passage has only been made by Wind class icebreakers.

The southern Beaufort Sea is usually passable from mid-June to mid-October. The passage around Point Barrow is only possible while the Arctic ice floes are held offshore by the summer southeast winds. Usually this condition prevails from early August to late September. The distribution of sea ice in the Chuckchi Sea and Bering Strait is controlled by the same winds. However, the ice retreats from these parts in late June and does not close them until October. The Bering Sea approach is navigable from early June to December.

Perhaps the most critical part of the passage is around Cape Barrow. Ships entering the Beaufort Sea must ensure their escape around the Cape, which may be closed by wind-drifted ice anytime after the first week of September. The approach to McClure Strait, along the west side of Banks Island closes at the same time. With these routes closed, the ships must be prepared to retreat eastward to the Atlantic Ocean before the passages through the Archipelago become frozen. Icebreakers can pass through Prince of Wales Strait to the eastward of Banks Island and traverse the shorter northern passage. Other ships must take the longer southern passage where the freeze-up occurs a few weeks later. The best route for surface vessels is from west to east by the southern passage. Most service to the DEW line radar stations has been by this route.

In 1958 the Northwest Passage was used as a submarine route when the nuclear powered submarine USS *Nautilus* made the first submarine passage from the Pacific to the Atlantic Ocean The first attempt in June was frustrated in the shallow (160-ft) Chuckchi Sea by an ice ridge which ex-



Northwest Passage, showing northern and southern routes. A west-to-east shipping route has been proposed from the open waters in the western Arctic, through Coronation Gulf, Dease, Victoria, Fury, and Hecla straits, into the Atlantic or Hudson Bay.

tended 80 ft below the surface. The remaining depth did not provide safe passage for the 50-ft height of the ship Returning through Bering Strait on July 29, the ship encountered the ice ridge [illegible] This time the ice ridge was avoided [illegible] trench provides a [illegible] the continental shelf *See* Arctic Ocean; Sea ice. [J.P.T.]

Nose

The structures surrounding or related to the nasal cavities. In man, the nasal cavities are triangular [illegible] that pass from the external openings, or [illegible] The [illegible] other. [illegible] por- tions of the ethmoid and [illegible] pro- jections of three turbinate bones on each side The floor of the nose is formed by the palate which is also the roof of the mouth The nasal cavities are lined with respiratory epithelium which also lines the paranasal sinuses The latter are cavities in the frontal, ethmoid, sphenoid, and maxillary bones, which communicate with the nasal passages *See* Cranium

The external nose consists of the nasal bones that form the bony bridge and two pairs of lower nasal cartilages These, together with the tightly adherent skin, determine the individual shape and size of each nose

Numerous blood vessels, nerves, and lymphatics supply and drain both the external and internal portions of the nose *See* Smell [E.G.ST.]

Nose cone

The forward portion of a missile or space vehicle that contains instruments and other payload The nose cone is required to withstand the operational conditions of launching and flight, plus the conditions encountered in reentering the earth's atmosphere (*see* Reentry) The nose cone decelerates because of drag as it penetrates the denser portions of the atmosphere, the kinetic energy of its motion

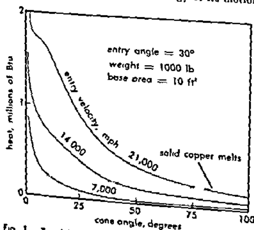
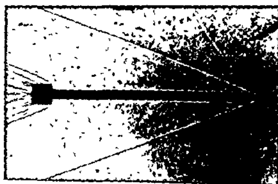


Fig 1 Total heating


Fig 2 Flow about slender body.

being transformed chiefly into heat. This heating of a nose cone has been likened to that encountered by meteors in the atmosphere. Most meteors tend to burn up before they reach the ground Their heating is a great deal more severe than that encountered by ballistic nose cones because meteors travel at a much higher average speed. Nevertheless, a nose cone can be expected to experience serious, if not destructive, heating; it is thus of foremost importance to study the nature of this heating

Amount of heating. The source of the heat transferred to a nose cone is primarily the friction work done on the vehicle by the air. This work is done adjacent to the vehicular surface in the boundary layer (*see* Ballistic range; Boundary-layer flow). The heating is higher in proportion to the amount of friction work done, and this work increases with the velocity. Both the amount of heat transferred to a nose cone and the rate at which it is transferred are important. The total amount of heat transferred is important because it determines the amount of material with which the nose cone must be provided to absorb this heat below a given temperature The rate at which heat is transferred is important because it determines the kind of material with which the nose cone must be provided to absorb the heat Figure 1 shows the total heating for cones of angles from 0° to 100°. Increasing cone angle means increasing bluntness of the body. The curves shown are the result of calculations assuming an entry angle between path and horizon of 30° for conical missiles weighing 1000 lb, having a base area of 10 ft² and entering the atmosphere at velocities of 7,000, 14 000, and 21,000 mph Total heating increases with increasing entry velocity, however, total heating decreases markedly with increasing cone angle or bluntness of the missile. For example, an entrance velocity of 14,000 mph, which is just a little less than that for the ICBM, requires at least a 40°-angle cone to prevent the cone from melting even if it is made of solid copper, and approximately a 100°-angle cone would be necessary to reduce the weight of copper required well below 50% of the total weight of the vehicle Figures 2 and 3 show why increasing the cone angle of the vehicle decreases the total heating These figures are photographs of flow about a slender and a blunt body, respectively In

Fig. 2 the boundary layer predominates in the flow about the slender body, and because this layer is the source of heating, it is not surprising that heating is high with a slender body. On the other hand, in Fig. 3, the shock wave predominates in the flow about the blunt body; this wave serves primarily to heat the air some distance from the body. As a rule of thumb, therefore, blunt bodies absorb less heat than slender bodies because most of their kinetic energy is delivered to the atmosphere in the form of heat, a relatively small amount appearing as heat in the body during entry. Accordingly, less of the weight of a blunt body is assigned to absorbing heat, and so more can be assigned to the payload.

Heating rate. If the heating rates are high, they cause the temperature to rise rapidly on the outer surface of the heat shield of a nose cone while it remains low on the inner surface. This may cause thermal deformations in the shell which are greater than it is structurally capable of withstanding, with the result that it may rupture during entry. Heating rates tend to be at a maximum at the nose of a

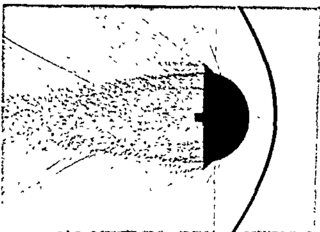


Fig 3. Flow about blunt body.

missile because this is where shearing, and hence heat generation, in the boundary layer occurs at the greatest rate This rate can be reduced by rounding the nose of the missile; the more nearly flat the nose, the less is the rate.

It is of interest to consider the same shapes that were treated previously in connection with total heating and to see what the gross effects of shape are on maximum heating rates. These effects are shown in Fig. 4 An increase in cone angle or over-all bluntness is favorable in the sense that it reduces heating rates. This reduction comes about because the blunt vehicles decelerate at higher altitudes where the air is less dense. From the standpoint of total heating and heating rates, blunt nose cones are especially attractive for atmosphere entry.

It is also true that blunt shapes can be made statically and dynamically stable during entry so that possibilities of tumbling in flight, leading to destruction of the missile, are minimized. On the other hand, blunt shapes tend to decelerate at rela-

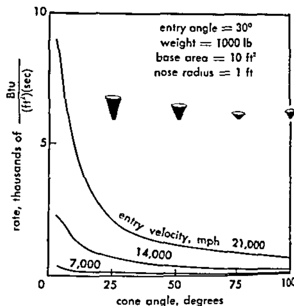
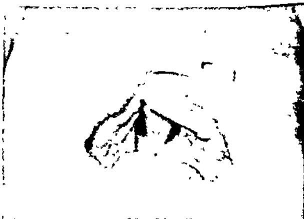


Fig 4. Maximum heating rate

tively high altitudes and therefore are especially susceptible to being drifted away from their target by cross winds. Slender shapes are less susceptible to wind drift and they tend, therefore, to be more accurate; because of the higher heating of these shapes, however, it is not possible to protect them during entry with a simple heat-shield material like copper Rather, it is necessary to employ heat shield materials with much higher heat capacity This fact has generated great interest in ablation materials which have heat capacities an order of magnitude greater than that of copper, due primarily to their high heats of vaporization Certain types of glass and plastics appear promising as ablation heat-shield materials for nose cones during atmosphere entry. *See* ASTRONAUTICS; GUIDED MISSILE; SPACECRAFT STRUCTURE. [A.J.EC.]

Nose disorders

Diseases of the nose include malformations, inflammatory processes, and, rarely, tumors. Among the malformations, a distortion of the nasal septum,



Bilateral cleft palate (cp) with unilateral cleft lip (cl) or harelip, extending into right nostril. Note the nasal septum in the depth of the cleft palate

Fig. 2 the boundary layer predominates in the flow about the slender body, and because this layer is the source of heating, it is not surprising that heating is high with a slender body. On the other hand, in Fig. 3, the shock wave predominates in the flow about the blunt body; this wave serves primarily to heat the air some distance from the body. As a rule of thumb, therefore, blunt bodies absorb less heat than slender bodies because most of their kinetic energy is delivered to the atmosphere in the form of heat, a relatively small amount appearing as heat in the body during entry. Accordingly, less of the weight of a blunt body is assigned to absorbing heat, and so more can be assigned to the payload.

Heating rate. If the heating rates are high, they cause the temperature to rise rapidly on the outer surface of the heat shield of a nose cone while it remains low on the inner surface. This may cause thermal deformations in the shell which are greater than it is structurally capable of withstanding, with the result that it may rupture during entry. Heating rates tend to be at a maximum at the nose of a missile because this is where shearing, and hence heat generation, in the boundary layer occurs at the greatest rate. This rate can be reduced by rounding the nose of the missile; the more nearly flat the nose, the less is the rate.

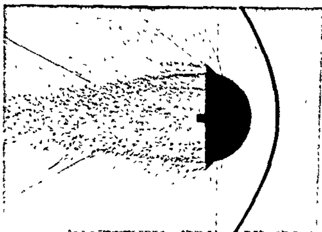


Fig 3. Flow about blunt body.

It is of interest to consider the same shapes that were treated previously in connection with total heating and to see what the gross effects of shape are on maximum heating rates. These effects are shown in Fig 4. An increase in cone angle or over-all bluntness is favorable in the sense that it reduces heating rates. This reduction comes about because the blunt vehicles decelerate at higher altitudes where the air is less dense. From the standpoint of total heating and heating rates, blunt nose cones are especially attractive for atmosphere entry.

It is also true that blunt shapes can be made statically and dynamically stable during entry so that possibilities of tumbling in flight, leading to destruction of the missile, are minimized. On the other hand, blunt shapes tend to decelerate at relatively high altitudes and therefore are especially susceptible to being drifted away from their target by cross winds. Slender shapes are less susceptible to wind drift and they tend, therefore, to be more accurate; because of the higher heating of these shapes, however, it is not possible to protect them during entry with a simple heat-shield material like copper. Rather, it is necessary to employ heat-shield materials with much higher heat capacity. This fact has generated great interest in ablation materials which have heat capacities an order of magnitude greater than that of copper, due primarily to their high heats of vaporization. Certain types of glass and plastics appear promising as ablation heat-shield materials for nose cones during atmosphere entry. *See* ASTRONAUTICS; GUIDED MISSILE; SPACECRAFT STRUCTURE. [A.J.EG]

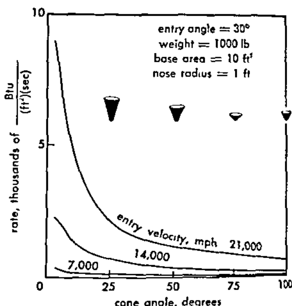
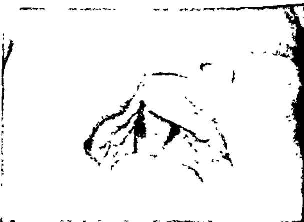


Fig 4. Maximum heating rate.

Nose disorders

Diseases of the nose include malformations, inflammatory processes, and, rarely, tumors. Among the malformations, a distortion of the nasal septum,



Bilateral cleft palate (cp) with unilateral cleft lip (cl) or harelip, extending into right nostril. Note the nasal septum in the depth of the cleft palate

Fig. 2 the boundary layer predominates in the flow about the slender body, and because this layer is the source of heating, it is not surprising that heating is high with a slender body. On the other hand, in Fig. 3, the shock wave predominates in the flow about the blunt body, this wave serves primarily to heat the air some distance from the body. As a rule of thumb, therefore, blunt bodies absorb less heat than slender bodies because most of their kinetic energy is delivered to the atmosphere in the form of heat, a relatively small amount appearing as heat in the body during entry. Accordingly, less of the weight of a blunt body is assigned to absorbing heat, and so more can be assigned to the payload.

Heating rate. If the heating rates are high, they cause the temperature to rise rapidly on the outer surface of the heat shield of a nose cone while it remains low on the inner surface. This may cause thermal deformations in the shell which are greater than it is structurally capable of withstanding, with the result that it may rupture during entry. Heating rates tend to be at a maximum at the nose of a

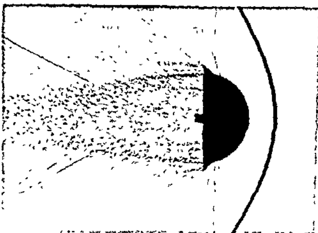


Fig 3 Flow about blunt body

missile because this is where shearing, and hence heat generation, in the boundary layer occurs at the greatest rate. This rate can be reduced by rounding the nose of the missile; the more nearly flat the nose, the less is the rate

It is of interest to consider the same shapes that were treated previously in connection with total heating and to see what the gross effects of shape are on maximum heating rates. These effects are shown in Fig. 4. An increase in cone angle or over-all bluntness is favorable in the sense that it reduces heating rates. This reduction comes about because the blunt vehicles decelerate at higher altitudes where the air is less dense. From the standpoint of total heating and heating rates, blunt nose cones are especially attractive for atmosphere entry.

It is also true that blunt shapes can be made statically and dynamically stable during entry so that possibilities of tumbling in flight, leading to destruction of the missile, are minimized. On the other hand, blunt shapes tend to decelerate at rela-

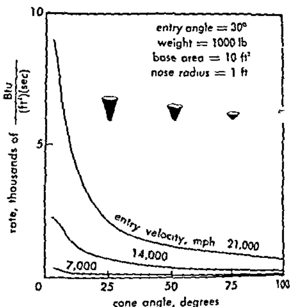
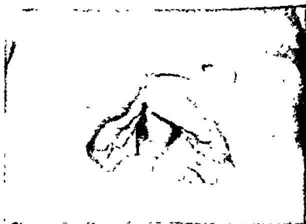


Fig. 4. Maximum heating rate.

tively high altitudes and therefore are especially susceptible to being drifted away from their target by cross winds. Slender shapes are less susceptible to wind drift and they tend, therefore, to be more accurate; because of the higher heating of these shapes, however, it is not possible to protect them during entry with a simple heat-shield material like copper. Rather, it is necessary to employ heat-shield materials with much higher heat capacity. This fact has generated great interest in ablation materials which have heat capacities an order of magnitude greater than that of copper, due primarily to their high heats of vaporization. Certain types of glass and plastics appear promising as ablation heat-shield materials for nose cones during atmosphere entry. See ASTRONAUTICS; GUIDED MISSILE; SPACECRAFT STRUCTURE. [A.J.EG]

Nose disorders

Diseases of the nose include malformations, inflammatory processes, and, rarely, tumors. Among the malformations, a distortion of the nasal septum,



Bilateral cleft palate (cp) with unilateral cleft lip (cl) or harelip, extending into right nostril. Note the nasal septum in the depth of the cleft palate

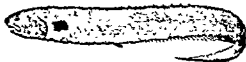
which can narrow one nasal duct to a considerable extent, is very common. An incomplete formation of the palate leaves a cleft (see illustration), through which the oral and nasal cavities communicate. It can be associated with a gap in the upper teeth and with a fissure in the upper lip which extends into the nostril resulting in hare lip. An infant with this malformation is able neither to suck nor to swallow properly. Later it will have severe speaking difficulties unless the malformation is corrected surgically.

Acute rhinitis is an inflammation of the mucous membrane of the nose due to either infection or allergy. In either case the swollen mucous membrane secretes abundant mucus, which is more viscid in infectious than in allergic rhinitis. The nasal ducts are completely obstructed. Mucus can be trapped in the paranasal sinus (see SINUSITIS). Infectious rhinitis, the common cold, is very often encountered in man but occurs in all animals. Allergic rhinitis is due to acquired hypersensitivity to various substances of the environment. These substances are frequently pollen or hay dust. See ALLERGY, ATOPIC.

Chronic rhinitis can be hyperplastic, if the tissues of the mucous membrane proliferate, and lead to nasal polyps. It can also be atrophic and result in gradual destruction of the mucous membrane, the glands and, finally, even the bone. A dry nose results. See ATROPHY. [E.WE.]

Notacanthiformes

The spiny eels form an order of actinopterygian fishes of medium size. This order is also known as the Heteromi. The body is elongated, tapers posteriorly and has no caudal fin. There is no duct to the swimbladder, the orbitosphenoid pterosphenoid opisthotic, and basisphenoid bones are absent, the posttemporal is simple or ligamentous, the transverse processes are not suturally joined to vertebral centra, there is no mesocoracoid arch, the pelvic fin is abdominal in position, with many rays, and the anal fin is long.



Spiny eel, *Notacanthus nasus*. (After D. S. Jordan and B. W. Evermann, The Fishes of North and Middle America, U.S. Natl. Museum Bull. 47, 1900)

This small order, which has a history extending back to the Upper Cretaceous, includes three families and six Recent genera. Spiny eels inhabit deep seas of all oceans, some have photophores. Of uncertain relationship, they are like true eels in that they lack a firm suspension of the pectoral girdle from the skull, but some have true fin spines like the perciform fishes. See ACTINOPTERYGII, PHOTOPHORE GLAND. [R.M.B.]

Notodelphyoida

A small group of crustaceans comprising several related families usually found within the body cavity of sedentary tunicates. Approximately 50 genera have been described from estuarine and marine waters throughout the world; no fresh-water representatives are known. The relationship between guest and host appears to be highly variable. Frequently, obvious signs of parasitism are lacking. The copepod usually resides as a commensal in the large sea-water-filled cavity of the tunicate and utilizes food gathered by the latter. In cases of obvious parasitism the species may exhibit extensive morphological degeneration while adapting to a specialized existence such as partial to complete encystment within the host's tissue.

Taxonomy. Recent systematic studies have raised serious questions regarding the validity of the order Notodelphyoida. In K. Lang's (1948) reorganization of the group, unrelated families were eliminated, the five acceptable families were divided into two subgroups, and phylogenetic placement of the group was handled in a manner most consistent with known facts. In agreement with several previous workers, Lang acknowledged the close kinship between notodelphyids and the Cyclopinidae, a family of the Cyclopoida, section Gnathostoma. He considered the two groups as distinct lines within the Gnathostoma ranking them as tribes, Notodelphyidiformes and Cyclopinidiformes. P. Illg (1958), reviewing the American notodelphyids, showed that Notodelphyoida as conceived by G. O. Sars is untenable. Illg revised the family Notodelphyidae by enlarging it to include doropygid and buprorid ...

us, ...

rule in adults ... dimorphism is the

o ...

r

a

f ... observed but the actual food utilized is unknown. For example, *Ascidicola rosea*, found in the transparent ascidian, *Corella parallelograma*, usually holds fast to the ... string ...

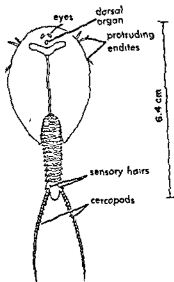
crosc

gut.

string, thus avoiding entry into the ... on the food

ho

So ... notodelphys), others



Lepidurus arcticus, female, dorsal aspect

ventral groove The young hatch in the nauplius larval stage

The group contains only two genera, *Triops*, or *Apus*, and *Lepidurus*, and about a dozen species. They live in nonpermanent waters, sometimes occurring in great numbers, and are distributed all over the world Their eggs are able to withstand extreme desiccation when the pools dry up. See BRANCHIOPODA [F L.]

Notoungulata

These dominant, hoofed herbivores of the Cenozoic of South America are abundantly represented in Paleoc-- ... tary re ... occurr ... early l ... primitive creodont ancestry at an early date, they radiated into a wide diversity of forms, some of which were convergent with Northern Hemisphere ungulates

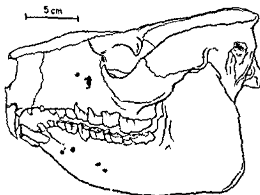


Fig 1 Skull and jaw of *Adinotherium ovinum*, an early Miocene toxodontid notoungulate from the Santa Cruz formation of Patagonia, Argentina (After W. Scott)

Notoungulates were characterized by a skull with expanded temporal region due to presence of a large sinus in the squamosal (Fig. 1) and no postorbital bar. The dentition is primitive with full eutherian formula and a tendency to retain a closed tooth row, although some groups enlarge the median incisors and develop a gap between the incisor row and cheek teeth by reduction of the posterior incisors, canines, and anterior premolars. There was always a complete molar row, and incomplete molarization of the posterior premolars. The teeth were low- to high-crowned, some groups developing high-crowned ever-growing teeth The cusps were joined by ridges in the upper and lower molars (lophodonty) even in the earliest forms. Characteristically, the upper molars had a straight ectoloph (Fig 2*a*), oblique protoloph and transverse metaloph, the median valley enclosed by the three lophs usually carried accessory cusps. The lower molars typically were crescentic, the protolophid and hypolophid resembling the Perissodactyla. The entoconid was isolated (Fig. 2*b*) or connected transversely to the hypoconid rather than to the hypoconulid.

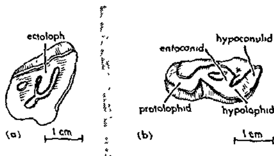
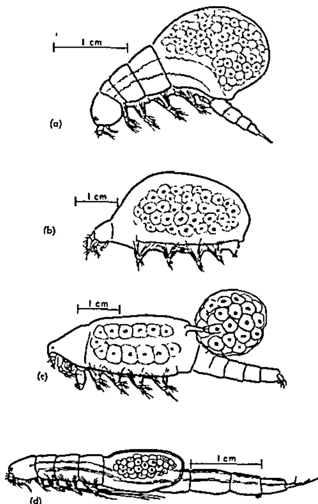


Fig 2 *Adinotherium ovinum*, from the Santa Cruz formation of Patagonia, Argentina (a) Upper molar (after W. Scott) (b) Lower molar. Note the typical crescentic shape

The feet were primitive with five toes, reduced to three or two in some advanced forms, the weight borne mainly by the third digit. See DENTITION

Notoungulates are represented in the earliest known ... specialized side line rather than a truly ancestral group, and by the earliest members of a central stock, suborder Toxodontia, which gave rise to most of the middle and late Tertiary forms. Surprisingly, at about the same time in Central Asia the most abundant mammalian herbivores were members of the Notioprogonia A single specimen records the presence of this group in the early Eocene of North America A late Mesozoic origin, either in Asia or South America, seems possible; although knowledge of late Mesozoic and Paleocene faunas of the world is too incomplete at present to rule out other possibilities.

In late Paleocene time South America was isolated from North America; and from that time



Notodelphyoida (a) Female *Doropygus psyllus* Thorell (after G O. Sars, 1921). (b) Female *Buprorus loveni* Thorell (after G O. Sars, 1921). (c) Female *Botryllophilus brevipes* Sars (after G. O. Sars, 1921) (d) Female *Ascidicola rosea* Thorell (after G O. Sars, 1921).

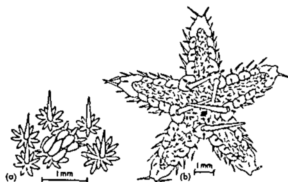
may choose the branchial chamber (*Doropygus*). In colonial ascidians the canals between zooids and the common exhalant pore, or cloaca, are frequent sites. Males usually retain the ability to leave the host and swim about, whereas females tend to become progressively more sedentary as they mature.

Reproduction. Eggs are retained in the female brood pouch until embryological development is completed When the young are ready to hatch the female expels them by arching her body. Several genera lack a thoracic incubatorium, having instead enlarged plates girdling a portion of the body One of these, *Ascidicola*, releases its young, still protected by an inner egg membrane, into the host's stomach The unhatched copepod passes through the intestinal tract protected by its egg membrane. At the anal opening it frees itself from the membrane and leaves the host as a free swimming larva. In general, larval development includes several naupliar instars with the infective stage occurring early in the copepodid phase of the life cycle. *See* ASCIDIACEA; CYCLOPOIDA. [A.F.]

Bibliography: R. V. Gotto, The biology of a commensal copepod, *Ascidicola rosea* Thorell, in the ascidian *Corella parallelogramma* (Müller), *J Marine Biol. Assoc. United Kingdom*, 36:281-290, 1957; P. L. Illg, North American copepods of the family Notodelphyidae, *Proc. U.S. Natl. Museum*, 107(3390):463–649, 1958; K. Lang, Copepoda "Notodelphyoida" from the Swedish west-coast with an outline on the systematics of the copepods, *Arkiv Zool.*, 40A(14):1–46, 1948; G. O Sars, *An Account of the Crustacea of Norway*, vol 8, 1921.

Notomyota

A suborder of Phanerozonia in which the upper marginals alternate in position with the lower marginals to impart a degree of flexibility to the arm, and each of the tube-feet has a terminal sucking disk Paxillae are present on the upper surface. Each arm usually contains a pair of dorsal muscles, whose contraction enables the arm to be turned upward over the disk These are mainly deep-water forms. *See* PHANEROZONIA. [H.B.F]



Representative Notomyota (a) *Cheiraster richardsoni*, paxillae and one pectinate pedicellaria (b) *Benthopecten pentacanthus*

Notostraca

An order of crustaceans of moderate size (20–90 mm) generally referred to the Branchiopoda They are called the tadpole-shrimps The cylindrical trunk consists of 25–44 body segments, and the number varies slightly within a species The first 11 somites each bear 1 pair of legs. Behind these are a varying number of segments with an increasing number of legs; a varying number of legless segments; and, finally, a telson with 2 narrow, cylindrical, caudal filaments The dorsal shield is rounded, and flattened dorsoventrally It usually covers only part of the animal The paired eyes are sessile and the antennules and antennae are much reduced The numerous legs, 35–71 in number, are of almost uniform structure They are flattened with marked-off endites. The genital openings are situated on the eleventh pair of legs, each of which, in the female, bears a circular brood pouch After the eleventh somite, the legs become smaller in size but increase in number independently of the number of somites. This is called polypody. The animals feed mainly on detritus, and the food is moved to the mouth in a

ar drops 3–4 magnitudes. Emission lines, especially of hydrogen and ionized iron, appear, as do multiple absorption lines.

Stage 7, the transition stage, is often marked by large oscillations of brightness and great complexity of the absorption spectrum. Bright lines increase and absorption lines disappear until the nebular spectrum predominates. For a short time, the star becomes a deep red because of the strength of the bright hydrogen alpha line.

In stage 8, the final decline, the brightness usually *decreases smoothly*. Absorption lines disappear, the continuum is weakened, and the spectrum becomes similar to that of a gaseous nebula. The color of the star becomes green because of the intense nebular lines. In stage 9, the postnova stage, many years after maximum, the star approaches its prenova state, and appears to be a blue dwarf. A nebular envelope may become visible, but the spectrum of the star is nearly continuous.

Spectra. The spectral changes of novae are complex and difficult to interpret. Several systems of absorption spectra may appear at one time, in addition to the nebular lines. The velocities of the various systems differ radically, and it is evident that many jets and shells of material are ejected from the stars at varying times and velocities.

Photographs of the brighter novae show jets or rings of nebulosity ejected from the central star image after the maximum stage. GK Persei (Nova 1901) was the first nova to show this phenomenon. Soon after the outburst, a nebulous shell was seen around the star. A series of photographs showed the shell expanding at a rate about equal to the velocity of light. It was assumed to be the illumination of an already existing nebulosity, by the sudden surge of light from the nova. Fifty years after the original outburst, photographs taken at the Palomar Observatory showed a reappearance of nebulosity with the same structure. This time it was probably the result of the shock waves produced in the diffuse nebulosity by the arrival of the ejected material from the nova.

A study of the structure of the lines in the spectrum of V603 Aquilae (Nova 1918) showed two large jets of material thrown out from the star about the time of maximum, one jet coming toward Earth, the other almost directly away from Earth. These, with a number of lesser jets, built a series of rings around the star. The material was moving outward with a velocity of 1700 km sec. About a year after maximum, photographs of the nova showed a small nebulous shell developing. By 1959, the nova was about eleventh ...

ar ...

at ...

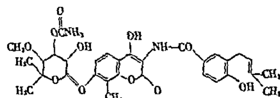
st ... nebulosity was complex, consisting of rings and jets or blobs of material.

In spite of the spectacular explosion, a nova actually loses only a small percentage of its total mass. The recurrent novae prove that a star can have several such explosions, and after each one return to a state similar to its prenova state.

Many theories have been advanced as to the cause of nova outburst. E. Schatzman considers a nova explosion a burst similar to a hydrogen bomb detonated by the helium He^3 atom. The shock wave could transfer almost all the energy to the surface, and some material would be ejected from the star. B. J. Bok believes that the internal sources of energy production deep inside the star become unstable and produce a giant internal explosion that shoots the star's atmosphere into space. [M W.M.]

Novobiocin

A moderately broad spectrum acid antibiotic first reported in 1955. Novobiocin is produced by strains of *Streptomyces niveus* and *Streptomyces spheroides*. In the Italian literature the same antibiotic is known as vulcamicina. It is a dibasic acid with the structural formula



Production. Biosynthesis is accomplished by deep-tank fermentation, distiller's solubles being an ingredient essential for high yields. The addition of ...

Novobiocin may be recovered from the filtered ...

... is formulated as capsules of the monosodium salt and liquid formulations of the ...

... in clinical use. Considerable protein binding takes place in ... serum ...

The ...

mate ...

hours ...

The antibiotic is relatively nontoxic unless it is given for more than 1 week. The most frequent side effect is a skin rash which may subside with continued therapy. The rash may or may not recur upon subsequent dosage.

Activity. Novobiocin is highly active against staphylococci and certain strains of *Proteus*. It is also active against pneumococci, streptococci, *Brucella*, and certain strains of *Escherichia* ...

until the Pliocene notoungulate evolution proceeded unharassed by competition from the herbivores of the Northern Hemisphere. A third suborder, the rodentlike Typotheria, appeared in the early Eocene. Toward the close of the Eocene the notoungulates displayed their greatest diversity and gave rise to a fourth suborder, the rodentlike Hegetotheria. Notioprogonians did not survive the Eocene. In the middle and late Tertiary the toxodonts, typotheres, and hegetotheres tended to specialize along rather restricted lines which resulted in a decrease in diversity within the order. The late Pliocene and Pleistocene emigrant ungulates from North America eventually replaced the large, hippopotamuslike toxodontids, bear-sized typotheres and rabbitlike hegetotheres in South America. Toxodontids spread north into Central America in the Pleistocene.

[R.H.T.]

Nova

A new star or a star that suddenly becomes bright as a result of an internal explosion. The term nova is misleading, because it is not in reality a new star, but is a brightening of an existing faint star. The objects might better be called spasmodic or capricious stars.

Novae have been observed to increase as much as 10–15 magnitudes in less than 1 day. Some decrease rapidly, but others remain bright for many months Eventually they return to their earlier magnitudes *See* VARIABLE STAR

Because of the sudden and unpredictable explosions of the novae, few of them have been completely observed. Of the 153 novae listed in the catalogs, only 22 had been identified in the prenova stage, and few more than that number have been *observed after their return to normal minimum* brightness.

Fast, slow, and recurrent types. A critical inspection of the light curves of novae shows a tendency to fall into two general classes, fast novae *and slow novae. Fast novae quickly rise to maxi*mum, remain near maximum for a short time, then fade to their premaximum brightness in a few years or less. The slow novae may take a month or more to reach maximum, and many years for the decrease For example, GK Persei (Nova 1901) increased from thirteenth magnitude to 0 magnitude is less than 4 days; 2 weeks later it had faded to fourth magnitude Irregular fluctuations then appeared with a gradual fading. It returned to thirteenth magnitude in about 11 years. In contrast to this, a slow nova, RT Serpentis (Nova 1909) took more than a month to rise from fourteenth to the eleventh magnitude, and remained at maximum for about 15 years. During the next 15 years, it dropped to about thirteenth, and by 1945 it had returned to fourteenth magnitude.

Six novae are known to have had several explosions, and are called recurrent novae. The observed cycles range from about 20 to 80 years The light curves of recurrent novae are of both fast and slow types, and seem to have no unique features.

The most recently observed maximum of a recurrent nova is that of RS Ophiuchi in July, 1958 Its previous outbursts were in 1898 and 1933. These stars might well be called repeating novae; the light curve is so nearly the same each time that it is reasonably safe to extrapolate for a sparsely observed maximum.

Absolute magnitude. It is extremely difficult to determine absolute magnitudes of novae with certainty. They are very distant, and their trigonometric parallaxes are often negative and have little significance. Determination of their distances from interstellar lines is uncertain because the distribution of absorbing material is not uniform. A possible method depends on the rate of expansion of the nebular disk correlated with radial velocities. The best values are obtained from the study of novae in other galaxies of known distances, such as the Magellanic Clouds and the Andromeda nebula

There is a correlation between the light curves and maximum brightness of novae wherever they are found, in our galaxy or other galaxies. More than 100 classical novae have been observed in the Andromeda nebula. They appear to be similar to the novae in our system, and comparisons may be made which aid the determination of absolute magnitudes.

An absolute magnitude of −7 6 at maximum has been adopted for the novae from the mean of many determinations by different methods. Before the explosion, the star probably is a blue subdwarf with absolute magnitude of about +4.

Galactic novae are concentrated in a band 10° each side of the plane of the galaxy and are densest toward the center of the galaxy. According to C Payne-Gaposchkin, 63% of the known novae lie *in the quadrant that contains the galactic center,* and 44% are between galactic longitudes 320 and 340°. In the Andromeda nebula, the same situation exists, and the novae are observed to be concentrated toward its center

Stages of a nova. *The spectra of normal galactic* novae undergo a series of similar changes during the development and decline of the brightness of the star D. B McLaughlin lists nine stages in the development of a nova In stage 1, prenova, the star is a blue dwarf and the spectrum is continuous. Stages 2 and 3 cover the initial rise to the frequently observed premaximum halt, about 2 magnitudes below the maximum brightness The star is expanding rapidly and the spectrum shows absorption lines displaced toward the violet Occasionally, strong emission lines appear at the time of the halt. Stage 4, the final rise, is much slower than the initial rise Absorption lines are displaced in the spectrum and the emission lines disappear. The greatly distended photosphere is similar to that of a supergiant star.

Stage 5, the principal maximum, is very brief except in the case of slow novae. The spectrum changes rapidly and emission lines appear in a few hours. The color of the star changes from blue to yellow. Stage 6, the early [illegible] lasts [illegible]

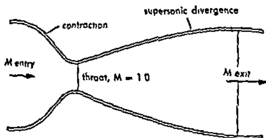


Fig 2 Variation of pressure, temperature, velocity and area with Mach number, Isentropic flow of air.

but, to keep pressure loss in the upstream ducting at a minimum, the entry Mach number should be low Contractions of 10 or greater are common, in which case the resulting entry Mach number is 0.06 (Fig. 2). Another advantage of large contractions is that they permit more effective use of screens to produce low turbulence

The ratio of exit to throat area of the supersonic section is fixed by the desired exit Mach number. To obtain a uniform, parallel exit flow free from shock disturbances, the divergent contour must be carefully designed The usual procedure requires two steps: first, the theoretical wall shape for nonviscous flow is obtained Then boundary layer corrections are added to obtain the final shape

Because the utility of a nozzle is greatly increased if its exit Mach number can be varied, the area ratio may be made adjustable by flexing or translating the divergent walls

Condensation. Experience shows that the large temperature drop in nozzles for high Mach numbers can result in condensation of one or more of the constituents of the working fluid Upstream dryers or heaters or both are usually employed to ensure a minimum temperature not more than 50°F below the condensation temperature. See GAS DYNAMICS [F D K]

Bibliography. H L. Dryden and T Von Karman (eds), *Advances in Applied Mechanics*, vol 5, 1958, A Ferri, *Elements of Aerodynamics of Supersonic Flows*, 1949, A H Shapiro, *Dynamics and Thermodynamics of Compressible Fluid Flow*, 2 vols., 1954.

Nuclear aircraft propulsion

The use of nuclear fuel as the energy source to move a vehicle Nuclear fission's 1,700,000-fold energy per pound advantage over petroleum fuel gives it a tremendous potential for propulsion. But this advantage, to be exploited, requires the development of new technologies

Aircraft application. Two basic designs for aircraft drive appear feasible, schematically termed the open air (Fig 1) and the closed liquid (Fig. 2) cycles Both cycles may use turbojet, turboprop, or ramjet engines. The turboprop design may be expanded to include a gas generator and separate propeller drive turbines. Multiple sets of turbomachinery per reactor may be used with either cycle Damaging neutron and gamma radiation resulting from the fission process is a principal consideration. See NUCLEAR RADIATION (BIOLOGY). Weight of shielding to contain this radiation makes achieving satisfactory thrust-weight ratios difficult.

Selection of an open air or a closed liquid cycle depends on whether the inherent advantages of using the environmental atmosphere as the only working fluid surpass the merits of a system in which the reactor size is smaller because of greater cooling fluid density. In the open cycle, a hot reactor heats the air as it flows through. Therefore, sufficient flow area must be allowed for the air to pass through the reactor and through the reactor shielding. This leads to a heavy system, because shield

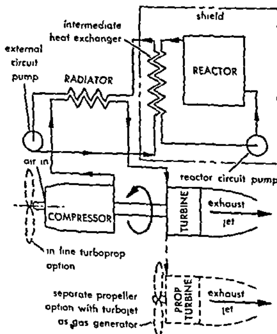


Fig 1. Open air nuclear engine cycle passes working fluid through reactor

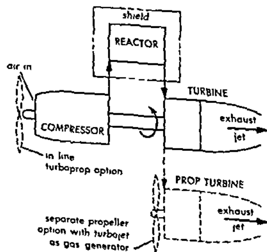


Fig. 2 Closed liquid nuclear engine cycle interposes a heat exchanging liquid between reactor and engine

clinically primarily as an antistaphylococcal drug, its annual production volume in 1958 reached approximately 15,000 kilograms. *See* AEROBACTER; ANTIBIOTIC; BRUCELLACEAE; ESCHERICHIA; PNEUMOCOCCUS; PROTEUS; PSEUDOMONAS AERUGINOSA; STAPHYLOCOCCUS. [C.M.S.]

Bibliography: M. Finland and R. Nichols, Novobiocin, *Antibiotica et Chemotherapia*, 4:209, 1957; C. G. Smith et al., Streptonivicin, a new antibiotic: I, discovery and biologic studies, *Antibiotics and Chemotherapy*, 6(2):135–142, 1956; L. S. Suter and E. W. Ulrich, Routine bacterial sensitivity studies, *Antibiotics and Chemotherapy*, 9(1): 38–46, 1959; H. Wallick et al., Discovery and antimicrobial properties of cathomycin, a new antibiotic produced by *Streptomyces spheroides*, n. sp., *Antibiotics Ann.*, 909–917, 1955–1956; H. Welch (ed.), *Antistaphylococcal Antibiotics*, in press.

Nozzle

A projecting opening that directs the flow of fluid into an open space. Some nozzles maintain the fluid in a jet; an example is the needle nozzle that directs water against the buckets of an impulse turbine. Other nozzles disperse the fluid in an atomized mist; an example is the cone nozzle that sprays liquid fuel into a combustion chamber. The nozzle may be an integral part of a machine, as the nozzle in a steam turbine, or it may be a separate interchangeable piece as on a fire truck.

Energy exchanges. The quantity Q of incompressible liquid such as water discharged from a smooth-walled nozzle supplied by liquid at head h at the entrance to the nozzle convergence is

$$Q = CA_2\sqrt{2gh}\sqrt{\frac{1}{1 - (A_2/A_1)^2}}$$

where A_1 is the entrance area at which head h is measured, A_2 is the discharge area, g is constant of gravity, and C is the discharge coefficient for the particular nozzle structure. A smooth tapered nozzle has a coefficient near 0.98; rough-walled nozzles and nozzles with abrupt changes in diameter have smaller coefficients, the coefficient being an indication of portion of the pressure head converted into discharge velocity.

In an atomizing nozzle, some of the pressure energy is expended in separating the liquid into droplets.

In a nozzle for compressible vapor or gas, as a steam nozzle, the energy changes are best determined by following the action through the nozzle on an enthalpy chart (*see* ORIFICE; STEAM). Experience indicates that actual velocity will be 0.98–0.96 the ideal velocity because of friction losses. For jet propulsion, the nozzle converts chamber pressure to exhaust velocity. High temperature combustion products may undergo dissociation, introducing a further energy exchange within the nozzle.

Wind-tunnel nozzle. As used in a wind tunnel, a nozzle increases fluid velocity but with the added requirement that the higher velocity stream be uniform and parallel. Physically the nozzle consists of a contracting section. If the final fluid velocity is to be supersonic, a divergent portion downstream of the contraction is also required (Fig. 1). The region at the minimum section is called the throat. The shape of the cross section is arbitrary; however, most tunnel nozzles are either circular (axisymmetric) or rectangular (2-dimensional).

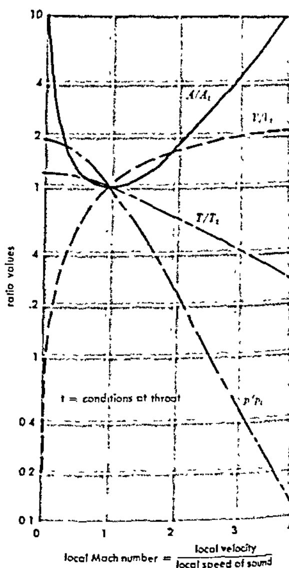


Fig. 1 Supersonic nozzle

Because the fluid, in passing through the nozzle, neither produces work nor gives up heat and because, in addition, area changes are usually gradual, the one-dimensional isentropic relations between fluid properties and velocity are useful approximations (*see* FLUID-FLOW PROPERTIES). Local pressure, temperature, flow area, and velocity relative to their values at the throat plotted as functions of local Mach number for air show that during acceleration the pressure and temperature decrease continuously; the area, however, must first decrease, then increase (Fig. 2).

Design considerations. In the contracting section the velocity should increase fairly uniformly and there should be no local regions of risin[illegible] sure. The amount of area [illegible] n is

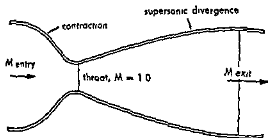


Fig. 2 Variation of pressure, temperature, velocity and area with Mach number, isentropic flow of air.

but, to keep pressure loss in the upstream ducting at a minimum, the entry Mach number should be low. Contractions of 10 or greater are common, in which case the resulting entry Mach number is 0 06 (Fig 2). Another advantage of large contractions is that they permit more effective use of

shock disturbances, the divergent contour must be carefully designed. The usual procedure requires two steps: first, the theoretical wall shape for nonviscous flow is obtained Then boundary layer corrections are added to obtain the final shape

Because the utility of a nozzle is greatly increased if its exit Mach number can be varied, the *area ratio* may be made adjustable by flexing or translating the divergent walls

Condensation. Experience shows that the large temperature drop in nozzles for high Mach numbers can result in condensation of one or more of the constituents of the working fluid Upstream dryers or heaters or both are usually employed to ensure a minimum temperature not more than 50°F below the condensation temperature. *See* GAS DYNAMICS [F D K]

Bibliography. H L Dryden and T Von Karman (eds), *Advances in Applied Mechanics*, vol 5, 1958; A Ferri, *Elements of Aerodynamics of Supersonic Flows*, 1949, A H Shapiro, *Dynamics and Thermodynamics of Compressible Fluid Flow*, 2 vols , 1954

Nuclear aircraft propulsion

The use of nuclear fuel as the energy source to move a vehicle Nuclear fission's 1,700,000-fold energy per pound advantage over petroleum fuel gives it a tremendous potential for propulsion But this advantage, to be exploited, requires the development of new technologies

Aircraft application. Two basic designs for aircraft drive appear feasible, schematically termed the open air (Fig 1) and the closed liquid (Fig 2) cycles. Both cycles may use *turbojet*, *turboprop*, or ramjet engines The turboprop design may be expanded to include a gas generator and separate propeller drive turbines Multiple sets of turbomachinery per reactor *may be* used with either cycle. Damaging neutron and gamma radiation resulting from the fission process is a principal consideration *See* NUCLEAR RADIATION (BIOLOGY). Weight of shielding to contain this radiation makes achieving satisfactory thrust-weight ratios difficult.

Selection of an open air or a closed liquid cycle depends on whether the inherent advantages of using the environmental atmosphere as the only working fluid surpass the merits of a system in which the reactor size is smaller because of greater cooling fluid density. In the open cycle, a hot reactor heats the air as it flows through. Therefore, sufficient flow area must be allowed for the air to pass through the reactor and through the reactor shielding. This leads to a heavy system, because shield

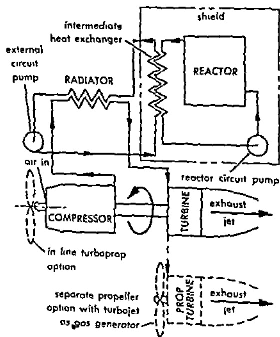


Fig 1. Open air nuclear engine cycle passes working fluid through reactor

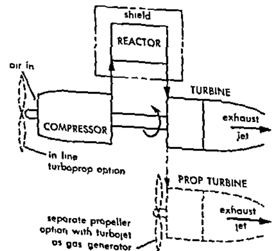


Fig 2. Closed liquid nuclear engine cycle interposes a heat exchanging liquid between reactor and engine

clinically primarily as an antistaphylococcal drug, its annual production volume in 1958 reached approximately 15,000 kilograms. *See* AEROBACTER; ANTIBIOTIC; BRUCELLACEAE; ESCHERICHIA; PNEUMOCOCCUS; PROTEUS; PSEUDOMONAS AERUGINOSA; STAPHYLOCOCCUS. [C.M.S.]

Bibliography: M. Finland and R. Nichols, Novobiocin, *Antibiotica et Chemotherapia*, 4:209, 1957; C G. Smith et al., Streptonivicin, a new antibiotic: I, discovery and biologic studies, *Antibiotics and Chemotherapy*, 6(2):135–142, 1956; L S. Suter and E. W. Ulrich, Routine bacterial sensitivity studies, *Antibiotics and Chemotherapy*, 9(1): 38–46, 1959; H. Wallick et al, Discovery and antimicrobial properties of cathomycin, a new antibiotic produced by *Streptomyces spheroides*, n. sp., *Antibiotics Ann.*, 909–917, 1955–1956; H. Welch (ed.), *Antistaphylococcal Antibiotics*, in press

Nozzle

A projecting opening that directs the flow of fluid into an open space Some nozzles maintain the fluid in a jet; an example is the needle nozzle that directs water against the buckets of an impulse turbine. Other nozzles disperse the fluid in an atomized mist, an example is the cone nozzle that sprays liquid fuel into a combustion chamber. The nozzle may be an integral part of a machine, as the nozzle in a steam turbine, or it may be a separate interchangeable piece as on a fire truck.

Energy exchanges. The quantity Q of incompressible liquid such as water discharged from a smooth-walled nozzle supplied by liquid at head h at the entrance to the nozzle convergence is

$$Q = CA_2\sqrt{2gh}\sqrt{\frac{1}{1-(A_2/A_1)^2}}$$

where A_1 is the entrance area at which head h is measured, A_2 is the discharge area, g is constant of gravity, and C is the discharge coefficient for the particular nozzle structure. A smooth tapered nozzle has a coefficient near 0.98; rough-walled nozzles and nozzles with abrupt changes in diameter have smaller coefficients, the coefficient being an indication of portion of the pressure head converted into discharge velocity.

In an atomizing nozzle, some of the pressure energy is expended in separating the liquid into droplets

In a nozzle for compressible vapor or gas, as a steam nozzle, the energy changes are best determined by following the action through the nozzle on an enthalpy chart (*see* ORIFICE; STEAM). Experience indicates that actual velocity will be 0 98–0.96 the ideal velocity because of friction losses. For jet propulsion, the nozzle converts chamber pressure to exhaust velocity. High temperature combustion products may undergo dissociation, introducing a further energy exchange within the nozzle

Wind-tunnel nozzle. As used in a wind tunnel, a nozzle increases fluid velocity but with the added

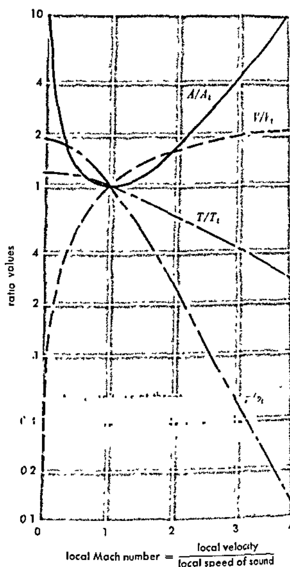


Fig 1. Supersonic nozzle.

requirement that the higher velocity stream be uniform and parallel Physically the nozzle consists of a contracting section If the final fluid velocity is to be supersonic, a divergent portion downstream of the contraction is also required (Fig. 1). The region at the minimum section is called the throat. The shape of the cross section is arbitrary; however, most tunnel nozzles are either circular (axisymmetric) or rectangular (2-dimensional)

Because the fluid, in passing through the nozzle, neither produces work nor gives up heat and because, in addition, area changes are usually gradual, the one-dimensional isentropic relations between fluid properties and velocity are useful approximations (*see* FLUID-FLOW PROPERTIES). Local pressure, temperature, flow area, and velocity relative to their values at the throat plotted as functions of local Mach number for air show that during acceleration the pressure and temperature decrease continuously; the area, however, must first decrease, then increase (Fig. 2).

Design considerations. In the contracting section the velocity should increase fairly uniformly and there should be no local regions of rising pressure. The amount of is

A recent high-voltage model, shown in Fig. 1, employs tritium gas sorbed in a thin layer of zirconium metal as the radioactive source. This source is looped around and spot welded to the center tube

sion The glass-insulated terminal is sealed to the nickel enclosure. The enclosure is evacuated through the center tube, which is then pinched off and sealed

The Radiation Research Corporation model R-1A is ¾ in in diameter and 0 531 in in height It weighs 0 2 oz and occupies 0 05 in³ It delivers about 500 volts at 160 micromicroamperes Future batteries are expected to deliver 1 microampere at

thousandth the toxicity of strontium-90. Both strontium-90 and krypton-85 require shielding to reduce external radiation to safe levels. Tritium produces no external radiation through a wall that is thick enough for any structural purpose Tritium was selected on the basis of low toxicity and no measurable external radiation

The principal use of the high-voltage battery is to maintain the voltage of a charged capacitor. The current output of the radioactive source is sufficient for this purpose

This type of battery may be considered as a constant-current generator The voltage is proportional to the load resistance. The current is determined by the number of emissions per second captured by the collector and does not depend on ambient conditions or the load As the isotope ages, the current declines For tritium, the intensity drops 50% in a 12 year interval For strontium-90 the intensity drops 50% in a 25 year interval

Low voltage nuclear batteries. *Three different* concepts have been employed in the low-voltage type of nuclear batteries. (1) a thermopile (2) the use of an ionized gas between two dissimilar metals, and (3) ...

into

into

T

voltage type employing a thermopile depends on the heat produced by radioactivity It has been calculated that a sphere of polonium-210 of 0 1 in diameter, which ...

suspended

surface te

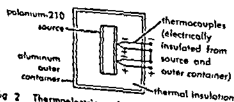


Fig 2 Thermoelectric nuclear battery.

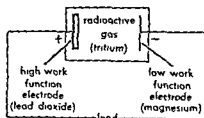


Fig. 3 Gas-ionization nuclear battery (schematic).

sivity of 0 25 For use as a heat source, it would have to be hermetically sealed in a strong, dense capsule. Its surface temperature, therefore, would be lower than 2200°C.

To complete the thermoelectric battery, the heat source must be thermally connected to a series of thermocouples which are alternately connected thermally, but not electrically, to the heat source and to the outer surface of the battery. After a short time, a steady-state temperature differential will be set up between the junctions at the heat source and the junctions at the outer surface This creates a voltage proportional to the temperature drop across the thermocouples

The battery voltage decreases as the age of the heat source increases With polonium-210 (half-life, 138 days) the voltage decreases about 0.5% each day With strontium-90, the decrease would be only about 0 01% per day (half-life, 20 years).

A battery containing 57 curies of polonium-210 sealed in a sphere 0 4 in. in diameter and 7 chromel-constantan thermocouples delivered a maximum power of 1 8 milliwatts It had an open-circuit voltage of 42 millivolts with a 78°C temperature differential. Over a 138 day period, the total electrical output would be about 1.5×10^4 joules (watt-sec).

Total weight of the battery was 34 g. This makes the energy output per pound equal to

$$\frac{1 5 \times 10^4}{3600} \text{ watt-hours (whr)} \times \frac{1}{34} \times 454 = 55 6$$

This is of the same magnitude as with conventional electric cells using chemical energy This nuclear energy, however, is being dissipated whether ...

On

210 se

... cost of a battery having an over-all efficiency of 0 2% has been estimated to be $3000 for the first milliwatt and $1600 per milliwatt increment. Reduced cost of this isotope and more efficient energy conversion will, of course, cut the cost of the battery proportionally.

The choice of isotope for a thermoelectric nuclear battery is somewhat restricted Those with a half life of less than 100 days would have a short useful life, and those with a half life of over 100 years would give too little heat to be useful This leaves 137 possible isotopes. This number is further reduced by the consideration of shielding.

The efficiency of the thermoelectric nuclear battery is dependent on the absolute temperature of

weight is largely determined by the volume to be shielded The large air ducts through the shield require extra material to counter the streaming effect. Another disadvantage is that compact reactors with large internal voids require large quantities of enriched uranium fuel to maintain the reaction. Although, in the open air cycle, higher than normal engine compressor pressure is used to increase the air density and decrease the required flow areas, thermodynamic limitations prevent achieving as small reactor size and shield weight as are permitted with the denser liquids of the closed cycle. Despite the liquid cycle's basic advantage of lower weight, the air cycle is receiving most of the initial development effort because it is a more practical method of using nuclear energy.

In a closed system, the liquid that recirculates through a reactor may become radioactive and require shielding. This may require a shielded intermediate heat exchanger in the liquid system so that the radioactive reactor liquid will not be carried to the relatively large engine radiators wherein the reactor heat is transmitted to the engine air. Application of the closed cycle is difficult because most suitable liquids solidify well above ambient temperature, are corrosive at the temperatures required for flight, and many are spontaneously inflammable in air with consequent fire hazard if a leak develops. However, because of the liquid system's basic advantage of smaller reactor size and lower shield weight, research is being supported to develop materials and designs wherein corrosion and other limitations are reduced to a tolerable level

Operational requirements. According to design studies conducted since 1946, range is the major predicted advantage of nuclear drive It will be difficult for nuclear-powered aircraft to achieve flight speeds and altitudes possible with shorter-range chemically fueled engines because of the weight of shielding necessary to provide sufficient protection to flight crews so that they will not be limited to a few flights per year. Since shield characteristics allow power-plant thrust-weight ratios to improve with increased thrust capacity, 300,000-pound gross weight or larger aircraft are required for nuclear drive. The existence of radiation will require extensive special handling facilities for servicing nuclear engines and airplanes. For these reasons nuclear drive is suitable only for applications where range capability is all important. Support of nuclear engine development is being maintained because it is the only present hope for manned aircraft with world-wide flight capability without using advanced base or in-flight refueling.

Space application. For propulsion beyond the earth's atmosphere, the propeller option is no longer available. All propulsion systems other than jets are in early conceptual stages compared to the preceding application to manned aircraft (see NUCLEAR ROCKET). In one system the working fluid is recirculated in a closed cycle for continuous operation. The turbine drives a generator to supply high voltage for electrical space propulsion (see ELECTROMAGNETIC PROPULSION; INTERPLANETARY PROPULSION). Such a drive system requires lightweight electrical generating equipment. The primary figure of merit for space power supplies is the specific weight, the power-plant weight per unit power (lb/kw). This parameter varies with power level, design life, state of the art, and meteoroid vulnerability. Nuclear closed cycle systems using gases like helium and argon or two phase systems with metals like sodium, rubidium, or potassium seem attractive. Heat generated by the reactor is delivered to the working fluid which is then expanded through a turbine as a vapor. An electric generator attached to the turbine shaft provides the electric power. Waste heat is rejected by radiation to space. Specific weights from 5 to 100 lb/kw have been calculated for turboelectric systems at the one-megawatt level. An important drawback of closed cycle systems is the vulnerability of the radiator to meteoroid penetration. Radiator walls must be made up to 0.100 in. thick to prevent excessive loss of working fluid.

Atomic recoil propulsion. Thrust can also be generated from nuclear action by expelling atomic fragments in a preferred direction The radioisotope sail is a typical example. If a large, thin, plastic film is coated on one side with a radioactive substance like polonium 210, the emergent alpha particles will be absorbed by the film on one side but will escape the sail on the other, generating thrust by recoil action. Accelerations of 10^{-5} g may be obtained (see RADIOACTIVITY).

[C.S.L.; W.R.C.]

Nuclear battery

A battery that converts the energy of particles emitted from atomic nuclei into electric energy Two basic types have been developed: (1) A high-voltage type, in which a beta-emitting isotope is separated from a collecting electrode by a vacuum or a solid dielectric, provides thousands of volts but the current is measured in micromicroamperes (10^{-12}); (2) a low-voltage type gives about 1 volt with current in microamperes (10^{-6}).

High-voltage nuclear battery. In the high-voltage type, a radioactive source is attached to one electrode, emitting charged particles. The source might be strontium-90, krypton-85 or hydrogen-3 (tritium), all of which are pure beta-emitters An adjacent electrode collects the emitted particles A vacuum or solid dielectric separates the source and the collector electrodes.

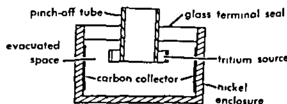


Fig 1. Tritium battery in cros on.

(Spontaneous transformations of unstable nuclei, such as α and β-decay or spontaneous fission, may be considered as a special type of nuclear reaction in which no external excitation by bombarding particles is involved.) Impurities in the target materials may also give rise to radioactive products which will contaminate the product of interest Chemical separations therefore are employed to obtain the product of interest free from the bulk of target material, from other reaction products, and from any radioactive contaminants which would interfere either with the determination of its yield in the reaction or with the study of its nuclear properties In the simple case of slow neutron capture in a very pure target, where only one reaction product is formed, the product is isotopic with the target, and it may be desirable to separate the product isotope from the bulk of the target material to increase the specific activity. This may also be true for the products of such nuclear reactions as (γ,n), $(n,2n)$ and (d,p) See NUCLEAR REACTION. Although isotopes are not usually separated in conventional chemical procedures, a special technique, the Szilard Chalmers process may be employed In this process, the recoil energy of the product of a nuclear reaction is sufficient to break its chemical binding in a molecule, and if its resultant chemical state is stable and does not interchange readily with inactive species present, it may be possible to separate it chemically from the target material

Technique of radiochemical analysis. The chemical manipulations and separations of radioactive materials, generally referred to as techniques of radiochemical analysis differ from ordinary analytical techniques in a number of respects Procedures in radiochemical analysis are not required to provide quantitative recovery, but are selected for specificity and speed, with reasonably good yields (usually the order of 50% or better) generally sufficing The criteria of radiochemical purity in a radioactive preparation are somewhat more stringent than those of ordinary chemical purity. Thus, trace quantities of impurities are rarely of importance in ordinary quantitative analyses, but in a radioactive preparation, contamination by trace quantities of radioactive impurities may completely negate the results of an experiment

The handling of highly radioactive materials presents a health hazard, and special techniques for the manipulation of samples behind shielding walls must be utilized Some effects of high levels of radioactivity on the solutions, such as heating and the decomposition of solvents which produces bubbling also may affect normal procedures

The mass of radioactive material produced in nuclear reactions is usually very small The concentrations of nuclear reaction products in the solutions of target materials are generally of the order of 10^{-10} M or less Many normal chemical operations such as precipitation, are not feasible with such small concentrations. Although separations can be carried out with these tracer quantities using such techniques as solvent extraction and ion exchange it then is difficult to determine the ... for the recovery of the product. Moreover

For example, radiocolloid formation, adsorption on the walls of vessels and on the surfaces of dust particles and precipitates, and the concentration dependence of some equilibrium constants become prominent at such extremely high dilution. To avoid these difficulties, an isotope dilution technique may be employed in which macroscopic quantities of the stable isotopes of the element of interest are added to serve as a carrier for the radioactive species

The amount of carrier used represents a compromise between considerations of convenience in chemical manipulations and yield determination, and the preparation of high specific activity sources in which counting corrections for absorption and scattering of radiations in the sample itself are minimized Quantities of 10–20 mg are used most often. Chemical procedures are simplified if macroscopic quantities of only a few elements are present When many elements are produced in a nuclear reaction (in nuclear fission, for example) aliquots of the solution usually are taken for the analysis of each element or small group of elements. It is then necessary to add carriers for only a relatively few products of interest. Tracer quantities of the other elements present are removed in the chemical procedures by the use of scavenging precipitations of a compound of high surface area [such as iron(III) hydroxide, or manganese dioxide], which tend to occlude traces of foreign substances, or of a representative precipitate for an insoluble group of elements, such as bismuth sulfide to carry trace quantities of other insoluble sulfides, lanthanum fluoride for insoluble fluorides, or iron(III) hydroxide for insoluble hydroxides. If the element of interest itself forms a precipitate which may occlude traces of other elements, it may be necessary to add hold-back carriers for the latter to dilute the effect of radioactive contamination of the product precipitate

For the isolat...

... the desired product, but which can be separated from it in the last stages of the procedure, leaving the product essentially carrier-free. Such carriers are referred to as nonisotopic carriers When it is necessary to determine the yield of a nuclear reaction product, a known quantity of an isotopic carrier must be used It is also imperative that complete interchange between the valence states of the carrier and the active species be achieved before any chemical separation is begun. In the case of elements which do not have any stable isotopes, or when a carrier-free procedure is desired, a known quantity of an isotopic radioactive tracer may be used Of course, the radiations of the tracer should be easily distinguishable from those of the product. The fractional recovery of the added carrier or tracer then will represent the yield of the product of interest

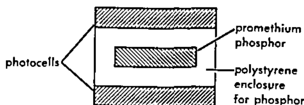


Fig. 4. Scintillator-photocell battery.

the heat source. The prototypes which have been reported had an efficiency of 0.1–0.2%.

The above discussion applies only to a portable power source. Drastic revision would be required if a thermoelectric-type nuclear battery were to be designed for central-station power.

Gas-ionization nuclear battery In this battery a beta-emitting isotope ionizes a gas situated in an electric field. Each beta particle produces about 200 current carriers (ions), so that a considerable current multiplication occurs compared with the rate of emission of the source. The electric field is obtained by the contact potential difference of a pair of electrodes, such as lead dioxide (high work function) and magnesium (low work function). The ions produced in the gas move under the influence of the electric field to produce a current.

A cell containing argon gas at 2 atmospheres, electrodes of lead dioxide and magnesium, and a radioactive source consisting of 1.5 millicuries of tritium has a volume of 0.01 in.^{3} and an active plate area of 0.2 in.^{2}, and gives a maximum current of 1.6×10^{-9} amp. The open-circuit voltage per cell depends on the contact potential of the electrode couple. A practical value appears to be about 1.5 volts Voltage of any value may be achieved by a series assembly of cells.

Scintillator-photocell nuclear battery. This type of cell is based on a two-step conversion process. Beta-particle energy is converted to light energy; then the light energy is converted into electric energy. To accomplish these conversions, the battery has two basic components, a light source and photocells.

The light source consists of a mixture of finely divided phosphor and promethium oxide (PM_2O_3) sealed in a transparent container of radiation-resistant plastic. The light source is in the form of a thin disk. The photocells are placed on both faces of the light source. These cells are modified solar cells of the diffused-silicon type.

Since the photocells are damaged by beta radiation, the transparent container of the light source must be designed to absorb any beta radiation not captured in the phosphor. Polystyrene makes an excellent light-source container because of its resistance to radiation

The light source must emit light in the range at which the photocell is most efficient A suitable phosphor for the silicon photocell is cadmium sulfide or a mixture of cadmium and zinc sulfide.

In a prototype battery, the light source consisted of 50 milligram[illegible] of the isotope [illegible] pure beta-emitt[illegible]

is deposited as a coating of hydroxide on the phosphor particles, which are then dried to give the oxide.

For a description of the photocell, *see* SOLAR BATTERY. For use with the low light level (about 0.001 times sunlight) of the light source, special treatment is necessary to make the equivalent shunt resistance of the cell not less than 100,000 ohms

The phototype battery, when new, delivers 20×10^{-6} amp at 1 volt. In 2.6 years (half-life) the current drops about 50% but the voltage drops only about 5%.

The power output improves with decreasing temperature, as a result of improved photocell diode characteristics which more than compensate for a decrease in short-circuit current. At −100°F, the power output is 1.7 times as great as at room temperature. At 144°F, the power output is only 0.6 times as great as at room temperature.

The battery requires shielding to reduce the weak gamma radiation to less than 9 milliroentgen per hour (mr/hr), which is the tolerance for continuous exposure of human extremities. The unshielded battery has a radiation level of 90 mr/hr By enclosing the cell in a case of tungsten alloy, density 16.5, the external radiation becomes less than 9 mr/hr

The unshielded battery has a volume of 0.014 in.^{3} and a weight of 0.016 oz. Over a 2.5 year period, the total output would be 0.32 whr (whether or not used). This gives a unit output of 320 whr/lb, which is about 6 times as great as chemical battery output

The shielded battery, however, has a volume of 0.07 in.^{3} and a weight of 0.6 oz. This reduces the unit output to 8.5 whr/lb

The cell can undergo prolonged storage at temperatures of 200°F. [S.El]

Nuclear chemistry

The study of the nuclear properties and reactions of nuclides through the use of chemical techniques for the isolation and purification of the species of interest. Nuclear chemistry is closely allied with nuclear physics, and the two fields are complementary. Although the chemical techniques employed in nuclear chemistry are essentially the same as those in radiochemistry, these fields of study may be distinguished on the basis of the aims of the investigations. Thus, a nuclear chemist utilizes chemical techniques as a tool for the study of nuclear reactions and properties, whereas a radiochemist utilizes the radioactive properties of certain substances as a tool for the study of chemical reactions and properties There is considerable overlap between the two fields, and in some cases (for example, the preparation and study of synthetic elements) a distinction may be difficult and somewhat arbitrary For a discussion of the applications of radioactive tracers to chemical problems *see* RADIOCHEMISTRY.

In a nuclear reaction, a bombarding particle interacts with a target nuclei [illegible] o produce more product nuclei and. [illegible]

(*see* FUSION, NUCLEAR; PARTICLE ACCELERATOR).
See also DECONTAMINATION (RADIOACTIVE CONTAMINANTS); PARTICLE DETECTOR; RADIOACTIVITY. [H.E.]

Bibliography: H Etherington (ed), *Nuclear Engineering Handbook*, 1958

Nuclear explosion

An explosion for which the energy is produced by a *nuclear transformation* (either fission or fusion)

The energy of a nuclear explosion is usually given *in terms of the equivalent energy release* of TNT using convenient units such as kiloton or megaton The prompt energy release of a nuclear explosion vaporizes the constituents and products as well as the bomb container These extremely hot gases are also at extremely high pressure and, following the explosion, exert enormous forces on their surrounding medium, such as air, earth, or water, thus initiating a complex series of effects which depend upon the surrounding medium Three types of bursts are distinguished: (1) air burst, (2) underground burst, and (3) underwater burst. See EXPLOSION AND EXPLOSIVE

Air burst. An air burst is defined as an explosion in the air above ground or water For a given energy release the exact effects of an air burst will depend upon the height of the explosion Nearly *all of the* prompt energy of the nuclear explosion, less the fraction of prompt γ rays and neutrons that escape the bomb, is coupled to the surrounding air in the form of an extremely strong shock wave which propagates outward, rendering the air luminous and creating a fireball in the *immediate vicinity* of the burst The shock wave propagates radially to *great distances* with decreasing strength If the explosion occurs close to the surface, there *will* be a *shock* wave coupled to the ground or water If the explosion is sufficiently close to *the surface*, a crater can be dug in ground, or a transient depression and waves created *in* water (Fig 1)



Fig 1. A nuclear detonation at medium height (air burst) beginning to evolve into characteristic mushroom shape

Table 1. Structural damage due to overpressure from 1-megaton air burst

Structure	Damage	Distance, miles
Glass windows, large and small	Shattering, occasional frame failure, 0 5–1 0 psi	10–20
Wood-frame building, one- or two-story houses	Wall framing cracked, roof badly damaged, interior partitions blown down	5 0–7 0
Multistory, wall-bearing building, brick apartment house, up to three stories	Exterior walls facing blast badly cracked, interior partitions badly cracked or down	3 5–4 2

Table 2. Structural damage due to drag loading from 1-megaton air burst

Structure	Damage	Distance, miles
Light-steel frame industrial building, one story, light walls	Some *distortion of* frame, windows *and* doors blown in, light siding ripped off	4–6
Medium-steel-frame industrial building, one story, 20-ton *crane* capacity, *light walls*	Some distortion of frame, windows and doors blown in, light siding ripped off	3–5
Multistory, steel-frame office-type building, five stories, low-strength walls	Moderate distortion of frame, interior partitions blown down, *all* windows and doors blown in	2 5–5

Blast effects The main material damage of an air-burst nuclear explosion is caused by the blast (shock) *wave* As the blast wave passes over a structure, differential pressures can cause severe damage to certain types of building Structures damaged by *overpressure* from a typical air burst [illegible] occur. At [illegible] distances the damage would be greater For other *energies*, similar damage would occur at distances proportional to the cube *root* of the energy.

The blast *wave also* imparts a material velocity to the air, and like the overpressure, this blast wave has its *maximum at the* shock front. The quantity of importance here is *the* dynamic pressure (one-half the air density in the shock wave times the square of the particle velocity). Certain types of structure are mainly *damaged by the* drag forces associated with the dynamic pressure Drag-sensitive structures and their damage distances are given in Table 2.

Window breakage and light structural damage can occur at abnormally large distances under certain meteorological conditions Large-yield explosions are sometimes *heard* at hundreds of miles and a 20-kiloton explosion has broken windows at 75–100 miles

Sample preparation and counting techniques. For studies in nuclear chemistry, the object of the radiochemical separations is the preparation of a pure sample in a form suitable for the radioactive assay of the nuclide of interest or for the determination of its nuclear properties. The detector used will, of course, depend on the type of radiation involved and the kind of information desired.

Alpha-particles and fission fragments have short ranges in matter, and to prevent absorption losses, samples of less than 100 $\mu g/cm^2$ surface density are generally required. A uniform sample deposit is necessary for accurate α-particle and fission-fragment measurements. This is accomplished most successfully by volatilizing, electroplating, or spraying on metal foils. The samples are counted internally in ionization chambers or proportional counters.

Beta-particles may cover a wide range of energies, and the techniques of sample preparation and counting will vary accordingly. The most commonly used detectors are Geiger, flow-type proportional, and scintillation counters. Samples may be prepared as indicated for α-emitters in the form of precipitates on filter paper disks or sample cups, as gases for internal counting, and as liquids. External sample counting usually is employed for convenience whenever feasible.

Gamma-radiation is highly penetrating, and the size or form of the sample is generally not very critical. Because of much higher efficiency, scintillation counters have essentially displaced all other detectors for γ-radiation.

Whenever possible, it is advisable to design experiments so that relative counting of samples will suffice. It is then necessary only to reproduce the counting conditions for each sample. The determination of absolute disintegration rates is a more difficult task, and many sources of error must be evaluated. These include such factors as the intrinsic efficiency of the detector, the fractional solid angle subtended by the detector at the source (usually called the geometry), the absorption and scattering of the radiation in the sample itself and in the material between it and the sensitive volume of the detector, and the backscattering of the radiation from the sample support. It is possible to eliminate or minimize the sources of error by the internal counting of samples in the form of a gas, in liquid scintillators, or in 4π-counters with thin samples mounted on thin supporting films. Beta-gamma coincidence counting also may be used when applicable. If none of these techniques is feasible or convenient, the counting conditions desired should be calibrated with sources of known disintegration rate, thus evaluating an overall conversion factor from observed activity to disintegration rate. When the highest accuracy is not paramount, the literature values for such correction factors may be used, with the counting conditions, of course, closely duplicating those reported.

Nuclear chemical investigations. The techniques of radiochemical analysis outlined above represent a powerful tool for the study of nuclear reactions and the properties of nuclides. They have led to the discovery of nuclear fission and of the naturally radioactive and synthetic elements, and to the elucidation of many aspects of complex nuclear reactions. Nuclear chemists have contributed much of the information on mass assignments, radiation types, half-lives, energies, disintegration schemes, and activation cross sections of the nuclides. The determination of excitation functions (that is, the energy dependence of a nuclear reaction) and the angular distribution of nuclear reaction products are other examples of nuclear chemical studies. In addition to these fundamental studies in nuclear science, many applied problems in the fields of isotope production, nuclear energy, and nuclear weapons fall in the realm of nuclear chemistry.

For a more detailed description of some of the topics referred to here *see* ISOTOPE; ISOTOPE SEPARATION (STABLE ISOTOPES); NUCLEAR STRUCTURE; PARTICLE DETECTOR; RADIOACTIVITY; TRANSURANIUM ELEMENTS. [E.P.S.]

Bibliography: G. B. Cook and J. F. Duncan, *Modern Radio-chemical Practice*, 1952; G. Friedlander and J. Kennedy, *Nuclear and Radiochemistry*, 1955.

Nuclear engineering

The branch of technology that deals with the utilization of the fission process. It is concerned with the design and construction of nuclear reactors and auxiliary facilities, the development and fabrication of special materials, and the handling and processing of reactor products. Development of the chain reaction for production of power and other purposes has required solution of difficult mechanical and metallurgical problems, and study of microscopic quantities of man-made elements to develop industrial processes for their chemical information. *See* FISSION, NUCLEAR; NUCLEAR POWER, REACTOR, NUCLEAR, REACTOR, NUCLEAR (CLASSIFICATION), REACTOR PHYSICS; REACTOR, SHIP PROPULSION.

For discussions of the means available for the protection of personnel from radiation, *see* HEALTH PHYSICS; MONITORING (IONIZING RADIATION); RADIATION SHIELDING. For discussions of the effects of radiation on materials, *see* RADIATION DAMAGE (INANIMATE MATERIALS). The handling and processing of irradiated fuel and other highly radioactive products of reactors are quite important (*see* NUCLEAR FUELS REPROCESSING, RADIOACTIVE WASTE DISPOSAL; RADIOCHEMICAL LABORATORY).

Radioisotopes and stable isotopes are produced for scientific and industrial use. *See* ISOTOPE SEPARATION (STABLE ISOTOPES), RADIOISOTOPE PRODUCTION. For discussions of the military aspects of the chain reaction, *see* ATOMIC BOMB, NUCLEAR EXPLOSION. Applications of engineering to processes not directly involving a usable product include design of particle accelerators and experimental equipment for developing fusion processes. Such topics are not here classed as nuclear engi-

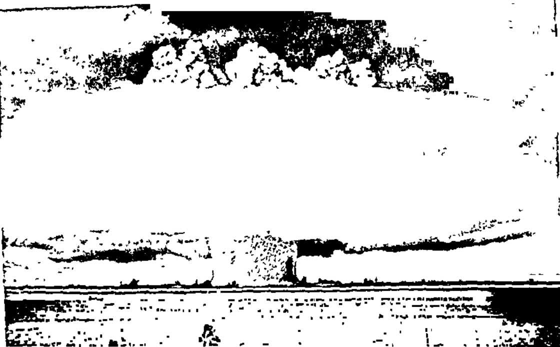


Fig 3 Column of water carried up by the venting bubble of an underwater explosion

bursts at the same distance from equal energy explosions

The gas bubble grows and collapses repeatedly and rises in the water Even for moderate depths no appreciable thermal radiation is measured on the surface even when the bubble vents the surface The radioactive fission products spread out along the surface of the water as the water column of the underwater explosion subsides (Fig 3) The radioactive fission debris becomes very dispersed in the water and weak in radiation intensity in about an hour [F H SH]

Detection of nuclear explosions. The various methods for detecting nuclear explosions depend on the detectable forms into which the energy emitted in such explosions is transformed (forms in which the signals may be transmitted over large distances) There are, however, natural phenomena which produce at the detection apparatus signals similar to those produced by nuclear explosions, and suppression of many of these signals from a nuclear explosion is possible by choice of environmental circumstances Identification of a nuclear explo[illegible]

on [illegible]

co[illegible]

teristic of a nuclear explosion

Among the effects which have been or could be employed for detection are the following:

1. Acoustic disturbances of low frequency ($^1/_{40}$ to 2 cps) resulting from the fireball in air Analogous hydroacoustic disturbances result from underwater explosions.

2. Radioac[illegible]

on a [illegible]

qua[illegible]

reasonable time of the event can be identified as resulting from a nuclear explosion by the ratios of fission products

3 Seismic disturbances, when enough energy couples into the earth. With a large enough signal, received at a number of seismic stations surrounding the source, many earthquakes can be identified as such, but nuclear explosions cannot at present be identified as such by their seismic signals

4 The generation of radio signals, through the asymmetric ionization produced in different directions by γ-radiation from the explosion.

5 Direct radiation of γ-rays, visible light, or thermal x-rays, which will be received by a detector which has a direct line of sight to the explosions, providing air or other absorbing media do not intervene (see Table 3)

The capabilities for detection and identification will thus differ with the medium in which the explosion takes place, as well as with the yield of the explosion. For explosions in the atmosphere, their acoustic waves, radio signals, and air-borne radioactivity will in general be useful, as will the seismic signal if the explosion is at low enough altitude Underwater and underground explosions produce hydroacoustic and seismic signals respectively, but no others; for identification of such events as natural or as nuclear explosions on-site inspection (to obtain a sample of debris if the event is nuclear) will be required, as it may with certain events in the lower atmosphere In the case of explosions at large distances from the earth, the direct radiations (or, in some cases, ionospheric disturbances which they may produce) would have to be employed, detection by instruments in satel-

Thermal effects. The light emitted from the luminous fireball causes thermal effects that depend upon the amount of thermal energy incident on exposed material or skin of man For a given air burst, the thermal energy per unit area decreases inversely as the square of the distance and, in addition, decreases because of atmospheric absorption and scattering. The intensity of thermal radiation at a given distance is proportional to the energy of the explosion.

Ignition of low-kindling materials, such as newspaper and rayon-acetate curtains, is caused by 3 cal/cm^2 of heat radiation and occurs at a radius of about 9 miles from an airburst of 1 megaton under good visibility conditions. At this same distance second-degree burns (characterized by painful blistering) would occur to exposed skin such as the face and arms. Considering that there are usually 20 exterior ignition points per acre of slum area and 3 per acre in a good residential area, fires and their subsequent spreading are likely in nearly all cities if exposed to a nuclear explosion.

Nuclear effects An air burst produces initial nuclear radiations of which the most harmful to man are the γ-rays and neutrons. At a distance of 1 mile from a 1-megaton explosion, the prompt γ-rays and neutrons would prove fatal to 50% of the exposed people (450 roentgens dose) even if they occupied a shelter with concrete walls 2 ft thick. This initial radiation dose decreases rapidly with distance from the burst; unsheltered personnel at 1.6 miles receive 450 roentgens (or equivalent), and, at 2 miles, only 30 roentgens. The dose at a given distance is roughly proportional to the energy of the explosion.

Fallout nuclear radiation becomes increasingly important as the relative height of the explosion decreases For a surface-burst 1-megaton explosion, 450 roentgens would be accumulated by a person in the open (unsheltered) during the first 18 hours following the explosion at a distance of about 130 miles downwind. Larger doses would be accumulated at nearer distances, and lesser doses at larger distances. Lethal fallout from a large-yield surface burst surpasses the range of all other effects, especially in the downwind direction

Radiation shelters. These would be most beneficial to large numbers of people, because the lethal fallout area can extend considerably beyond that for any other effect Effective fallout shelters have been designed and instrumented at various nuclear test sites. Some protection against fallout radiation is also afforded by structures such as homes (first floor of frame house reduced radiation dose by a factor of 2, and the basement of such a house by a factor of 10) and large buildings (multistory, reinforced concrete building reduced radiation dose by 10 for occupancy of the first floor and by 1000 or more in the basement). Protection against fallout nuclear radiation presents a host of difficult and involved problems: the radiation is invisible and requires special instruments for detection; its persistence requires occupation of the shelter for days or a week, depending upon the contamination level.

Underground burst. A nuclear explosion underground creates a sphere of extremely hot gases at high pressure consisting of vaporized earth and bomb which initiates a shock wave in the earth If the explosion takes place at relatively small depths, the shock wave vents the surface and large masses of rock and earth are carried into the air by the venting gases (Fig. 2). A shallow burst of 1 megaton in dry soil will produce a crater about 1250 ft in diameter and about 150 ft deep For other energies the diameters vary as the cube root of the energy and the depths approximately as the fourth root of the energy.

As the depth of the burst increases, both the amount of thermal radiation and the prompt nuclear radiation as measured on the ground surface decrease rapidly. The radioactive fallout is usually more intense but less widespread than for a surface burst

As the depth of the burst becomes large the explosion can be contained, and essentially no effects are observed on the earth's surface. A 1.7-kiloton explosion in Nevada located 900 ft vertically below a flat-top mountain was easily contained. In the particular rock formation (tuff) for this explosion a cavity about 55 ft in radius was formed inside a fused radioactive shell of earth about 10 cm thick surrounded by a thoroughly crushed zone of rock out to about 130 ft radius. The cavity soon collapsed because of the weight of the crushed rock For other energy explosions, the dimensions should scale as the cube root of the energy. The shock wave in the ground was recorded on seismographs hundreds of miles away from the 1.7-kiloton explosion.

Underwater burst. An underwater nuclear explosion forms a high-pressure steam bubble which initiates a strong shock wave propagating outward in the water Generally, the overpressures in water from underwater bursts are several orders of magnitude higher than the overpressures in air from air



Fig. 2. Shallow underground burst

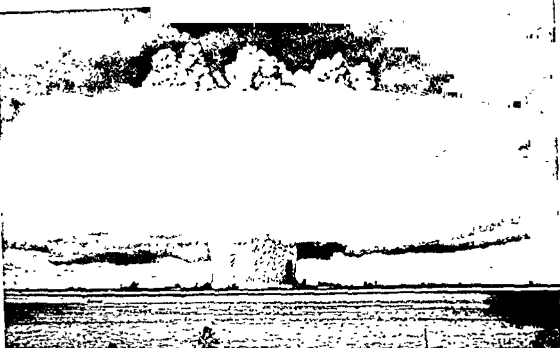


Fig 3 Column of water carried up by the venting bubble of an underwater explosion

bursts at the same distance from equal energy explosions

The gas bubble grows and collapses repeatedly and rises in the water Even for moderate depths no appreciable thermal radiation is measured on the surface, even when the bubble vents the surface The radioactive fission products spread out along the surface of the water as the water column of the underwater explosion subsides (Fig 3) The radioactive fission debris becomes very dispersed in the water and weak in radiation intensity in about an hour [F H SH]

Detection of nuclear explosions The various methods for detecting nuclear explosions depend on the detectable forms into which the energy emitted in such explosions is transformed (forms in which the signals may be transmitted over large distances) There are, however, natural phenomena which produce at the detection apparatus signals similar to those produced by nuclear explosions, and suppression of many of these signals from a nuclear explosion is possible by choice of environmental circumstances Identification of a [illegible]

[illegible]

to 2 cps) resulting from the fireball in air. Analogous hydroacoustic disturbances result from underwater explosions.

2. Radioactive [illegible]

reasonable time of the event can be identified as resulting from a nuclear explosion by the ratios of fission products.

3 Seismic disturbances, when enough energy couples into the earth With a large enough signal, received at a number of seismic stations surrounding the source, many earthquakes can be identified as such, but nuclear explosions cannot at present be identified as such by their seismic signals

4 The generation of radio signals, through the asymmetric ionization produced in different directions by γ-radiation from the explosion

5. Direct radiation of γ-rays, visible light, or thermal x rays, which will be received by a detector which has a direct line of sight to the explosions, providing air or other absorbing media do not intervene (see Table 3).

The capabilities for detection and identification will thus differ with the medium in which the explosion takes place, as well as with the yield of the explosion For explosions in the atmosphere, their acoustic waves, radio signals, and air borne radioactivity will in general be useful, as will the seismic signal if the explosion is at low enough altitude. Underwater and underground explosions produce hydroacoustic and seismic signals respectively, but no others; for identification of such events as natural or as nuclear explosions on-site inspection (to obtain a sample of debris if the event is nuclear) will be required, as it may with certain events in the lower atmosphere. In the case of explosions at large distances from the earth, the direct radiations (or, in some cases, ionospheric disturbances which they may produce) would have to be employed, detection by instruments in satel-

Thermal effects. The light emitted from the luminous fireball causes thermal effects that depend upon the amount of thermal energy incident on exposed material or skin of man. For a given air burst, the thermal energy per unit area decreases inversely as the square of the distance and, in addition, decreases because of atmospheric absorption and scattering The intensity of thermal radiation at a given distance is proportional to the energy of the explosion.

Ignition of low-kindling materials, such as newspaper and rayon-acetate curtains, is caused by 3 cal/cm^2 of heat radiation and occurs at a radius of about 9 miles from an airburst of 1 megaton under good visibility conditions. At this same distance second-degree burns (characterized by painful blistering) would occur to exposed skin such as the face and arms. Considering that there are usually 20 exterior ignition points per acre of slum area and 3 per acre in a good residential area, fires and their subsequent spreading are likely in nearly all cities if exposed to a nuclear explosion.

Nuclear effects An air burst produces initial nuclear radiations of which the most harmful to man are the γ-rays and neutrons At a distance of 1 mile from a 1-megaton explosion, the prompt γ-rays and neutrons would prove fatal to 50% of the exposed people (450 roentgens dose) even if they occupied a shelter with concrete walls 2 ft thick This initial radiation dose decreases rapidly with distance from the burst; unsheltered personnel at 1.6 miles receive 450 roentgens (or equivalent), and, at 2 miles, only 30 roentgens The dose at a given distance is roughly proportional to the energy of the explosion

Fallout nuclear radiation becomes increasingly important as the relative height of the explosion decreases. For a surface-burst 1-megaton explosion, 450 roentgens would be accumulated by a person in the open (unsheltered) during the first 18 hours following the explosion at a distance of about 130 miles downwind. Larger doses would be accumulated at nearer distances, and lesser doses at larger distances. Lethal fallout from a large-yield surface burst surpasses the range of all other effects, especially in the downwind direction

Radiation shelters. These would be most beneficial to large numbers of people, because the lethal fallout area can extend considerably beyond that for any other effect Effective fallout shelters have been designed and instrumented at various nuclear test sites. Some protection against fallout radiation is also afforded by structures such as homes (first floor of frame house reduced radiation dose by a factor of 2, and the basement of such a house by a factor of 10) and large buildings (multistory, reinforced-concrete building reduced radiation dose by 10 for occupancy of the first floor and by 1000 or more in the basement). Protection against fallout nuclear radiation presents a host of difficult and involved problems: the radiation is invisible and requires special instruments for detection; its persistence requires occupation of the shelter for days or a week, depending upon the contamination level.

Underground burst. A nuclear explosion underground creates a sphere of extremely hot gases at high pressure consisting of vaporized earth and bomb which initiates a shock wave in the earth If the explosion takes place at relatively small depths, the shock wave vents the surface and large masses of rock and earth are carried into the air by the venting gases (Fig 2). A shallow burst of 1 megaton in dry soil will produce a crater about 1250 ft in diameter and about 150 ft deep For other energies the diameters vary as the cube root of the energy and the depths approximately as the fourth root of the energy.

As the depth of the burst increases, both the amount of thermal radiation and the prompt nuclear radiation as measured on the ground surface decrease rapidly. The radioactive fallout is usually more intense but less widespread than for a surface burst.

As the depth of the burst becomes large the explosion can be contained, and essentially no effects are observed on the earth's surface. A 1.7-kiloton explosion in Nevada located 900 ft vertically below a flat-top mountain was easily contained. In the particular rock formation (tuff) for this explosion a cavity about 55 ft in radius was formed inside a fused radioactive shell of earth about 10 cm thick surrounded by a thoroughly crushed zone of rock out to about 130 ft radius The cavity soon collapsed because of the weight of the crushed rock For other energy explosions, the dimensions should scale as the cube root of the energy The shock wave in the ground was recorded on seismographs hundreds of miles away from the 1.7-kiloton explosion

Underwater burst. An underwater nuclear explosion forms a high-pressure steam bubble which initiates a strong shock wave propagating outward in the water. Generally, the overpressures in water from underwater bursts are several orders of magnitude higher than the overpressures in air from air



Fig. 2. Shallow underground burst.

Table 1. Nuclear reactions in a thermal-neutron spectrum

Reaction number	Equation	Cross section, barns*	Half-life
(1)	$U^{235} + n \rightarrow U^{236} + \gamma$	107	
(2)	$U^{235} + n \rightarrow$ Fission $+ 2.47\ n$	582	
(3)	$U^{238} + n \rightarrow U^{239} + \gamma$	2.74	
(4)	$U^{239} \rightarrow Np^{239} + \beta^-$		23.5 m
(5)	$Np^{239} \rightarrow Pu^{239} + \beta^-$		2.33 d
(6)	$Pu^{239} + n \rightarrow Pu^{240} + \gamma$	277	
(7)	$Pu^{239} + n \rightarrow$ Fission $+ 2.88\ n$	748	
(8)	$Pu^{240} + n \rightarrow Pu^{241} + \gamma$	250	
(9)	$Pu^{241} + n \rightarrow Pu^{242} + \gamma$	390	
(10)	$Pu^{241} + n \rightarrow$ Fission $+ 3.06\ n$	1025	
(11)	$Pu^{242} + n \rightarrow Pu^{243} + \gamma$	19	
(12)	$Pu^{243} \rightarrow Am^{243} + \beta^-$		4.98 h
(13)	$Th^{232} + n \rightarrow Th^{233} + \gamma$	7.3	
(14)	$Th^{233} \rightarrow Pa^{233} + \beta^-$		23.3 m
(15)	$Pa^{233} \rightarrow U^{233} + \beta^-$		27.4 d
(16)	$U^{233} + n \rightarrow U^{234} + \gamma$	52	
(17)	$U^{233} + n \rightarrow$ Fission $+ 2.51\ n$	527	
(18)	$U^{234} + n \rightarrow U^{235} + \gamma$	90	

* Accepted values for monoenergetic thermal neutrons at 2200 m/sec (0.0252 ev) 1 barn = 10^{-24} cm^2

clear fuel) of the system is given by the relative probability that reaction (3) will take place as compared to reactions (1) and (2)

Similarly, in a mixture of Pu^{239} and U^{238}, the conversion ratio (breeding ratio) is given by the relative probability of reaction (3) as compared to (6) and (7) In this case, however, the higher [illegible]

(9) The Pu^{243} formed by reaction (11) decays rapidly to americium, as shown, to end the chain effectively Thus after long exposure to thermal neutrons in a reactor, a mixture of U^{235} and U^{238} will contain appreciable concentrations of U^{236}, Pu^{239} Pu^{240}, Pu^{241} and Pu^{242}, all of which must be taken into consideration in determining the over all conversion ratio

Reactions (1) (2), and (13)–(18) represent the reactions taking place in a mixture of U^{235} and thorium In this fuel cycle secondary isotopes of importance to the conversion ratio are U^{233}, U^{234} U^{236} and Pa^{233} For most efficient neutron utilization (capture in thorium), it is important to minimize losses due to neutron absorption in Pa^{233}, by keeping the average neutron flux as low as possible

Maximizing conversion ratio When it is desired to maximize the neutron-conversion ratio in a reactor neutron losses are held to a minimum by suitable selection of the materials comprising the reactor system their arrangement in the reactor, and its operating conditions For example neutron leakage is reduced if the reactor is made large, fission product poisons (neutron absorbers) can be lowered by frequent processing of fuel; and nonfission neutron capture by fuel can be minimized by designing the reactor so that the average energy of the neutrons is optimum for causing fission However, the extent to which these methods of improving neutron utilization can be applied is limited by economic considerations. Thus, for any given nuclear power application, there will be an optimum reactor size and configuration and optimum fuel-processing cycle

The control of neutron losses due to parasitic capture in fuel by varying the relative amounts of neutron-scattering material (moderator) and fuel to give the proper neutron energy is the most important factor in achieving a high conversion ratio. The effect of neutron energy ν on α (the ratio of neutrons lost by parasitic capture in fuel to those leading to fission) and the number of neutrons emitted per neutron absorbed in fuel

$$\eta = \frac{\nu}{1 + \alpha}$$

is shown in Table 2.

Table 2 indicates that the theoretical maximum conversion ratio (given by $\eta - 1$) is above 1.0 for all three fissionable materials as long as the average energy of the neutrons causing fission is either very high (~1 Mev) or very low (~0.025 ev). In a practical reactor design, however, both of these neutron energy conditions are difficult to achieve, because for any given mixture of fuel and moderator there will exist neutrons moving at all energies, ranging from those for fission neutrons (fast or high-energy neutrons) down to those moving at approximately the same velocities as the moderator atoms Even in a highly thermalized reactor (high ratio of moderator to fuel), the neutron energy will vary considerably from the mean that is established by the moderator temperature. Because of this, the conversion ratio is affected by the moderator temperature This is especially true in the case of the U^{235}, U^{238}, Pu^{239} fuel cycle, as shown in Table 3.

In nuclear power reactors that operate with high moderator and coolant temperatures to achieve high thermal efficiencies, it is difficult to get a high con-

Table 2. Capture-to-fission ratio (α) and neutron yield (η) as functions of energy

Neutron energy, ev	U^{233}		U^{235}		Pu^{239}	
	α	η	α	η	α	η
0.025	0.102	2.28	0.190	2.07	0.380	2.09
0.10	0.08	2.33	0.17	2.11	0.59	1.81
0.30	0.15	2.19	0.25	1.97	0.70	1.70
10^3			0.52	1.62	0.72	1.67
10^5			0.18	2.09	0.60	1.80
10^6	0.03	2.44	0.08	2.28	0.10	2.62

Table 3 Effect of moderator temperature on the nuclear properties of U^{233} and Pu^{239}

Average moderator temperature, °C	Average neutron energy (kT), ev	Fast neutrons produced per thermal neutron absorbed in	
		U^{233}	Pu^{239}
75	0.030	2.083	2.006
200	0.041	2.094	1.936
350	0.054	2.102	1.875
600	0.075	2.103	1.871

Table 3. Methods available for detecting nuclear tests*

Method of detection	Type of test detected	Limitations on usefulness
Radioactive debris	Underground, underwater, surface and lower atm, possibly upper atm (≥50,000 ft)	Underground test requires on-site inspection
Air-acoustic	Surface and lower atm, possibly upper atm (≥50,000 ft)	Periodic windstorms
Hydroacoustic	Underwater	Earthquake background
Seismic	Underwater, underground, possibly surface and lower atm	Earthquake background
Electromagnetic waves	Surface and lower atm, upper atm (≥50,000 ft), possibly outer space	Severe natural-disturbance background
Visible light	Possibly surface and lower atm, upper atm (≥50,000 ft), possibly outer space	Periodic cloud cover, attenuation in air (for low-level tests)
X-rays	Possibly surface and lower atm, upper atm (≥50,000 ft), *outer space*	Can be shielded
Neutrons, γ-rays	Upper atm, (≥50,000 ft), outer space	Spread out in time (low intensity)

* From J C Mark, The detection of nuclear explosions, *Nucleonics*, 17(8) 64–73, 1959

lites being in general most useful More definite information on natural radiation backgrounds in space, and their variation with time, will allow a more precise estimate of distances at which such explosions can be detected and the reliability with which an identification can be made. *See* ATOMIC BOMB, FISSION, NUCLEAR; FUSION, NUCLEAR, HYDROGEN BOMB; RADIATION SHIELDING; RADIOACTIVE FALLOUT. [H.BN.]

Bibliography: S. Glasstone (ed), *The Effects of Nuclear Weapons*, 1957; G. W. Johnson et al., *The Underground Nuclear Detonation of September 19, 1957, Rainier Operation PLUMBBOB*, USDC UCRL-5124, 1958; J. C Mark, The detection of nuclear explosions, *Nucleonics*, 17(8):64–73, 1959.

Nuclear fuels

The fissionable and fertile elements and isotopes used as the sources of energy in nuclear reactors Although many heavy elements can be made to fission by bombardment with high-energy α-particles, protons, deuterons, or neutrons, only neutrons can provide a self-sustaining reaction

The number of neutrons ν released in the fission process varies from one per many fissions for elements just beyond the fission point (silver) to two or more per fission for the heavier elements, such as thorium and uranium. Even in such elements, neutron capture by the nucleus accompanied by th release of excess energy in the form of a γ-ra occurs in many cases, rather than nuclear fission This reduces the number of neutrons available fo further fission. The ratio of neutron capture t neutron fission varies from nucleus to nucleus an changes with the energy of the bombarding neu trons. Only a few isotopes of the heavy element have a higher probability of fission than captur These fissionable isotopes, U^{233}, U^{235}, and Pu^{239} are the only materials that can sustain the fissio reaction, and are therefore called nuclear fuel *See* REACTOR, NUCLEAR.

Of these isotopes, only U^{235} occurs in nature a 1 part in 140 of natural uranium, the remainde being U^{238}. The other two fissionable isotopes mu be produced artificially, U^{233} by neutron capture i Th^{232} and Pu^{239} by neutron capture in U^{238}. Th isotopes Th^{232} and U^{238} are called fertile ma terials.

By using a mixture of both fissionable and ferti isotopes in a nuclear reactor, it is possible to r duce the rate of depletion of the nuclear fuel b cause capture of excess neutrons by the fertile ma terial replenishes the fissionable material. Thu U^{235} can be burned (fissioned) and the surpl neutrons used to produce plutonium from U^{238} U^{233} from thorium Reactors in which such nucle processes take place are called converter reactor

The efficiency of production of new nuclear fu depends on the extent of neutron losses due undesirable neutron absorptions in the reactor to neutron leakage In some cases, these losses ca be kept small enough so that more nuclear fu is produced than burned. Moreover, the U^{233} ar Pu^{239} can be subsequently used as fuel in place the original U^{235}, and by this means a large fra tion of fertile material can be gradually convert into fissionable material Reactors that burn U and Pu^{239} and produce as much fuel as is co sumed, or more, are called breeders

The total energy that can be produced from th fissionable U^{235} in known resources of high-grad uranium ores corresponds to less than 5% of th from economically recoverable fossil fuels Thu atomic energy will not become an important sour of power unless the breeding and conversion fu cycles are utilized.

Breeding and conversion. The nuclear reactio governing the consumption and production of n clear fuel in a reactor are listed in Table 1. Al shown are values for the thermal-neutron (0 025- neutron) cross section (probability that the rea tion will take place) and the half-life for radi active decay of the relatively unstable isotopes

In a mixture of U^{235} and U^{238}, three competing reactions take place with thermal neutrons: (1) U^{235} capture, (2) U^{235} fission, and (3) U^{238} capture (numbers in parentheses refer to reactions in Table 1). Reaction (3) leads to the production of Pu^{239} by successive decay of U^{239} and Np^{239}, as shown by reactions (4) and (5). The conversion ratio (relative production and ... ion of nu-

Table 1. Nuclear reactions in a thermal-neutron spectrum

Reaction number	Equation	Cross section, barns*	Half-life
(1)	$U^{235} + n \rightarrow U^{236} + \gamma$	107	
(2)	$U^{235} + n \rightarrow$ Fission $+ 2.47\,n$	582	
(3)	$U^{238} + n \rightarrow U^{239} + \gamma$	2.74	
(4)	$U^{239} \rightarrow Np^{239} + \beta^-$		23.5 m
(5)	$Np^{239} \rightarrow Pu^{239} + \beta^-$		2.33 d
(6)	$Pu^{239} + n \rightarrow Pu^{240} + \gamma$	277	
(7)	$Pu^{239} + n \rightarrow$ Fission $+ 2.88\,n$	748	
(8)	$Pu^{240} + n \rightarrow Pu^{241} + \gamma$	250	
(9)	$Pu^{241} + n \rightarrow Pu^{242} + \gamma$	390	
(10)	$Pu^{241} + n \rightarrow$ Fission $+ 3.06\,n$	1025	
(11)	$Pu^{242} + n \rightarrow Pu^{243} + \gamma$	19	
(12)	$Pu^{243} \rightarrow Am^{243} + \beta^-$		4.98 h
(13)	$Th^{232} + n \rightarrow Th^{233} + \gamma$	7.3	
(14)	$Th^{233} \rightarrow Pa^{233} + \beta^-$		23.3 m
(15)	$Pa^{233} \rightarrow U^{233} + \beta^-$		27.4 d
(16)	$U^{233} + n \rightarrow U^{234} + \gamma$	52	
(17)	$U^{233} + n \rightarrow$ Fission $+ 2.51\,n$	527	
(18)	$U^{234} + n \rightarrow U^{235} + \gamma$	90	

* Accepted values for monoenergetic thermal neutrons at 2200 m/sec (0.0252 ev). 1 barn = 10^{-24} cm^2

clear fuel) of the system is given by the relative probability that reaction (3) will take place as compared to reactions (1) and (2).

Similarly, in a mixture of Pu^{239} and U^{238}, the conversion ratio (breeding ratio) is given by the relative probability of reaction (3) as compared to (6) and (7). In this case, however, the higher isotopes of Pu^{239} that are formed have a long half-life and start to absorb neutrons as their concentration builds up by means of reactions (6), (8), and (9). The Pu^{243} formed by reaction (11) decays rapidly to americium as shown, to end the chain effectively. Thus, after long exposure to thermal neutrons in a reactor, a mixture of U^{235} and U^{238} will contain appreciable concentrations of U^{236}, Pu^{239}, Pu^{240}, Pu^{241} and Pu^{242}, all of which must be taken into consideration in determining the over-all conversion ratio.

Reactions (1), (2), and (13)–(18) represent the reactions taking place in a mixture of U^{235} and thorium. In this fuel cycle, secondary isotopes of importance to the conversion ratio are U^{233}, U^{234}, U^{236}, and Pa^{233}. For most efficient neutron utilization (capture in thorium), it is important to minimize losses due to neutron absorption in Pa^{233}, by keeping the average neutron flux as low as possible.

Maximizing conversion ratio. When it is desired to maximize the ne[illegible] a reactor [illegible] suitable se[illegible] reactor sy[illegible] ...al arrangement in the reactor, and its operating conditions. For example, neutron leakage is reduced if the reactor is made large, fission product poisons (neutron absorbers) can be lowered by frequent processing of fuel; and nonfission neutron capture by fuel can be minimized by designing the reactor so that the average energy of the neutrons is optimum for causing fission. However, the extent to which these methods of improving neutron utilization can be applied is limited by economic considerations. Thus, for any given nuclear power application, there will be an optimum reactor size and configuration and optimum fuel-processing cycle.

The control of neutron losses due to parasitic capture in fuel by varying the relative amounts of neutron-scattering material (moderator) and fuel to give the proper neutron energy is the most important factor in achieving a high conversion ratio. The effect of neutron energy ν on α (the ratio of neutrons lost by parasitic capture in fuel to those leading to fission) and the number of neutrons emitted per neutron absorbed in fuel

$$\eta = \frac{\nu}{1 + \alpha}$$

is shown in Table 2.

Table 2 indicates that the theoretical maximum conversion ratio (given by $\eta - 1$) is above 1.0 for all three fissionable materials as long as the average energy of the neutrons causing fission is either very high (~1 Mev) or very low (~0.025 ev). In a practical reactor design, however, both of these neutron energy conditions are difficult to achieve, because for any given mixture of fuel and moderator there will exist neutrons moving at all energies, ranging from those for fission neutrons (fast or high-energy neutrons) down to those moving at approximately the same velocities as the moderator atoms. Even in a highly thermalized reactor (high ratio of moderator to fuel), the neutron energy will vary considerably from the mean that is established by the moderator temperature. Because of this, the conversion ratio is affected by the moderator temperature. This is especially true in the case of the U^{235}, U^{238}, Pu^{239} fuel cycle, as shown in Table 3.

In nuclear power reactors that operate with high moderator and coolant temperatures to achieve high thermal efficiencies, it is difficult to get a high con-

Table 2. Capture-to-fission ratio (α) and neutron yield (η) as functions of energy

Neutron energy, ev	U^{233}		U^{235}		Pu^{239}	
	α	η	α	η	α	η
0.025	0.102	2.28	0.190	2.07	0.380	2.09
0.10	0.08	2.33	0.17	2.11	0.59	1.81
0.30	0.15	2.19	0.25	1.97	0.70	1.70
10^3			0.52	1.62	0.72	1.67
10^5			0.18	2.09	0.60	1.80
10^6	0.03	2.44	0.08	2.28	0.10	2.62

Table 3. Effect of moderator temperature on the nuclear properties of U^{235} and Pu^{239}

Average moderator temperature, °C	Average neutron energy (kT), ev	Fast neutrons produced per thermal neutron absorbed in	
		U^{235}	Pu^{239}
75	0.030	2.083	2.006
200	0.041	2.091	1.936
350	0.054	2.102	1.875
600	0.075	2.103	1.871

version ratio because of the effect just described. One solution to this problem is to insulate the moderator thermally from the coolant and to maintain the moderator at a lower temperature. Such a technique cannot be applied in graphite-moderated reactors because of the necessity for keeping the graphite hot to minimize its expansion due to radiation damage and to minimize the buildup of stored energy. *See* RADIATION DAMAGE (INANIMATE MATERIALS).

Fast reactors. By eliminating the moderator, it is possible to raise the average neutron energy in a reactor to a value close to that of the fission neutrons (2.0 Mev, average). Coolants, fertile material, and structural material in the core, however, tend to degrade the energy so that the average is normally 0.6–0.2 Mev. Under these conditions, the ratio of parasitic fuel captures to fuel fissions varies from 0.12 to 0.25. Corresponding breeding ratios range from 1.96 to 1.40 for Pu^{239}-fueled fast (unmoderated) reactors and from 1.34 to 1.08 for U^{235}-fueled reactors. In all cases, the neutron yield from fissions in fuel is increased by fast-neutron fissions in fertile material (U^{238} or Th^{232}) resulting in a higher breeding ratio than that given simply by $\eta - 1$.

It is evident from the foregoing that considerably higher conversion or breeding ratios are possible in a U^{235}- or Pu^{239}-fueled fast reactor than in a thermal reactor. Fast reactors, therefore, provide a means of utilizing a far greater proportion of natural uranium than would be otherwise possible.

In the case of thorium utilization by means of the U^{233}-thorium cycle, breeding is possible with both fast and thermal neutrons. Here, the difference in breeding ratio between thermal and fast reactors is not as great as for the U^{235}-plutonium cycle, and the choice depends upon other considerations, such as the amount of fissionable material required for criticality in each case.

In addition to achieving a high conversion ratio in a nuclear power reactor, it is also desirable to have a high thermal efficiency and high material economy (heat output per unit weight of fuel and fertile material). Unfortunately, in most cases these three characteristics cannot be maximized simultaneously. For example, in a boiling water reactor, which generates steam inside the reactor core for power production, an increase in the rate of steam generation increases the neutron losses and decreases the neutron economy Therefore, the optimum design of this and most other reactor types involves a compromise between high power density and high neutron economy.

Fuel requirements. Estimates of the growth rate of the nuclear power system in the United States vary widely, depending upon the degree of optimism assumed. The most optimistic estimate predicts an installed nuclear plant capacity of 225,000,000 kilowatts by 1980, which is about twice the 1955 total electrical capacity in the United States. Less optimistic estimates indicate an installed nuclear capacity of only 35,000,000 kilowatts by 1980. By taking a geometric average of these extremes as a basis, it is possible to obtain an order of magnitude estimate of the long-range nuclear source material requirements. Data are summarized in Table 4, assuming also that the nuclear power system will be made up of equal numbers in terms of Mw(e) of converter reactors, fast breeders, and thermal breeders

Table 4. Cumulative nuclear source material requirements

	Year			
	1970	1980	1990	2000
Nuclear electric capacity, Mw(e)	11,000	88,000	270,000	600,000
Natural uranium, short tons				
Thermal converters	2,700	25,000	88,200	231,000
Fast breeders	2,300	14,000	21,000	
Thorium breeders	100	500	800	1 000
Total	5,100	39,500	110,000	232,000
Thorium, short tons	1,100	8,800	27,000	60,000

The inventory requirements per Mw(e) are 0 5 tons natural uranium or equivalent (converters); 4 8 kg U^{235} or Pu^{239} plus 0 6 tons U^{235} (fast breeders); 1.1 kg U^{235} or U^{233} plus 0 3 tons thorium (thorium breeders). The fuel consumption or production per MwY(e) is 0 06 tons natural uranium (converters), 0 5 kg Pu^{239} (fast breeders); 0 16 kg U^{233} (thorium breeders)

Reserves of uranium and thorium. Known reserves of high-grade ores from which uranium and thorium can be recovered at $10/lb (of oxide) or less are summarized in Table 5

Table 5. Known reserves of high-grade uranium and thorium

Country	Mineral	Lb per ton of ore	Short tons
Uranium, U_3O_8			
United States	Carnotite, autunite	5	230,000
Canada	Brannerite	2	400,000
South Africa	Pitchblende (in gold ore)	½	400,000
France			25,000
Other	Davidite, pitchblende		20,000
Total			1,075,000
Thorium, ThO_2			
United States	Thorite, monazite	65	20,000
Canada	Brannerite	1	200,000
India	Monazite	200	150,000
Brazil	Monazite	130	10,000
Other			20,000
Total			400,000

* After concentration

Estimates of additional reserves of high-grade uranium ores based on general geologic data and the discovery experience since 1950 indicate that the total uranium reserves in non-Communist countries may contain as much as 4,000,000 tons of U_3O_8. Lower-grade ores such as bituminous shale and phosphate deposits, [illegible] in ½ [illegible] of uranium per ton, are [illegible]

ranium reserve of more than 20,000,000 tons. ranium from these latter reserves, however, may st \$30-50/lb of uranium oxide in high-grade ncentrates

The uranium ore reserves of the Blind River and ancroft areas in Canada probably contain about 00,000 tons of thorium The cost of recovering is thorium as a by-product of uranium production ay be in the range of \$1-2/lb Less is known about otential reserves of higher-cost thorium; however, at some price, presumably many times the resent price, thorium availability may approach hat of uranium

It is evident from a comparison of the requirements and reserves of nuclear source materials that here is an adequate amount of high-grade ore to upport the nuclear power industry through the ear 2000, if the industry expands at a rate corresponding to an average of the various projections and nuclear plant types are distributed evenly between converters, fast breeders, and thorium breeders *See* RADIOACTIVE MINERALS

Preparation of uranium fuel. Starting with ore, four steps are required in the preparation of a natural uranium fuel These are (1) recovery of uranium from ore (concentration), (2) purification of crude conc...

following step (2) are (3) conversion of oxide to UF_6, (4) isotope separation by gaseous diffusion, (5) reduction of enriched UF_6 to metal, alloy, or compound, and (6) fabrication of the fuel element These steps are described as follows.

Concentration Because of the variety of natural sources of uranium, no one concentration method is uniquely suited to all ores. Concentration by gravity methods, for example, is applicable for

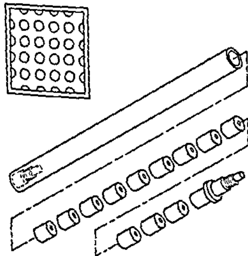


Fig 1. Cylindrical pellets of UO_2 are pressed to exacting specifications for size and weight After finishing, pellets are inserted into stainless steel or zircaloy tubes Tubes are sealed and welded, then assembled into bundles to form the rod-type element (From Nuclear Fuel Elements, General Electric)

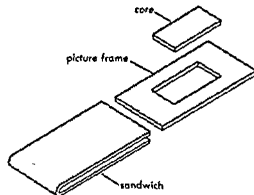


Fig 2. A fuel plate is assembled with the core, or uranium alloy piece, fitting into a picture frame of aluminum plate. Aluminum plate is then placed on either side and the entire sandwich is hot-rolled to effect bonding *After centering the core by x-ray, the plates are trimmed to size, assembled, and mechanically bonded to the side plates to form finished elements.* (From Nuclear Fuel Elements, General Electric)

pitchblende but not for carnotite or autunite, from which uranium is extracted almost exclusively by leaching with acid or alkali carbonate. This is followed by a precipitation process (or more recently by ion exchange or solvent extraction) to recover the uranium from the leach solutions.

Purification. To make natural uranium most suitable for use in a nuclear reactor, it is desirable to reduce the concentration of neutron-absorbing impurities such as boron, cadmium, and the rare earths to levels of 0 1-10 parts per million. This is accomplished either by selective extraction of uranyl nitrate from aqueous solutions by certain oxygenated organic solvents, notably diethyl ether, methyl isobutyl ketone, or tributyl phosphate in kerosine, or by quantitative precipitation of uranium peroxide ($UO_4 \cdot 2H_2O$) from weakly ...

... produce UO_3. This is reduced to UO_2, which is treated with HF to produce green salt, UF_4 Uranium metal is produced from green salt by reduction with calcium or magnesium metal and UF_6 gas is produced from UF_4 by reaction with fluorine.

Isotope separation Separation of the uranium isotopes, U^{235} and U^{238}, depends upon the physical differences arising from the difference in their atomic weights Gaseous diffusion is now used to take advantage of this difference. *See* ISOTOPE SEPARATION (STABLE ISOTOPES)

UF_6 reduction and fabrication. The UF_6 product from the diffusion plant must be reduced to uranium oxide or uranium metal for incorporation in fuel elements. For most reactor applications the fuel elem...

a

ni

must be in intimate contact with the uranium-bear... This cladding ing material ...

version ratio because of the effect just described. One solution to this problem is to insulate the moderator thermally from the coolant and to maintain the moderator at a lower temperature. Such a technique cannot be applied in graphite-moderated reactors because of the necessity for keeping the graphite hot to minimize its expansion due to radiation damage and to minimize the buildup of stored energy. *See* RADIATION DAMAGE (INANIMATE MATERIALS).

Fast reactors. By eliminating the moderator, it is possible to raise the average neutron energy in a reactor to a value close to that of the fission neutrons (2.0 Mev, average). Coolants, fertile material, and structural material in the core, however, tend to degrade the energy so that the average is normally 0.6–0.2 Mev. Under these conditions, the ratio of parasitic fuel captures to fuel fissions varies from 0.12 to 0.25 Corresponding breeding ratios range from 1.96 to 1.40 for Pu^{239}-fueled fast (unmoderated) reactors and from 1.34 to 1.08 for U^{235}-fueled reactors. In all cases, the neutron yield from fissions in fuel is increased by fast-neutron fissions in fertile material (U^{238} or Th^{232}) resulting in a higher breeding ratio than that given simply by $\eta - 1$.

It is evident from the foregoing that considerably higher conversion or breeding ratios are possible in a U^{235}- or Pu^{239}-fueled fast reactor than in a thermal reactor. Fast reactors, therefore, provide a means of utilizing a far greater proportion of natural uranium than would be otherwise possible.

In the case of thorium utilization by means of the U^{233}-thorium cycle, breeding is possible with both fast and thermal neutrons. Here, the difference in breeding ratio between thermal and fast reactors is not as great as for the U^{235}-plutonium cycle, and the choice depends upon other considerations, such as the amount of fissionable material required for criticality in each case.

In addition to achieving a high conversion ratio in a nuclear power reactor, it is also desirable to have a high thermal efficiency and high material economy (heat output per unit weight of fuel and fertile material). Unfortunately, in most cases these three characteristics cannot be maximized simultaneously For example, in a boiling water reactor, which generates steam inside the reactor core for power production, an increase in the rate of steam generation increases the neutron losses and decreases the neutron economy. Therefore, the optimum design of this and most other reactor types involves a compromise between high power density and high neutron economy.

Fuel requirements. Estimates of the growth rate of the nuclear power system in the United States vary widely, depending upon the degree of optimism assumed. The most optimistic estimate predicts an installed nuclear plant capacity of 225,000,000 kilowatts by 1980, which is about twice the 1955 total electrical capacity in the United States. Less optimistic estimates indicate an installed nuclear capacity of only 35,000,000 kilowatts by 1980. By taking a geometric average of these extremes as a basis, it is possible to obtain an order of magnitude estimate of the long-range nuclear source material requirements. Data are summarized in Table 4, assuming also that the nuclear power system will be made up of equal numbers in terms of Mw(e) of converter reactors, fast breeders, and thermal breeders

Table 4. Cumulative nuclear source material requirements

	Year			
	1970	1980	1990	2000
Nuclear electric capacity, Mw(e)	11,000	88,000	270,000	600,000
Natural uranium, short tons				
Thermal converters	2,700	25,000	88,200	231,000
Fast breeders	2,300	14,000	21,000	
Thorium breeders	100	500	800	1,000
Total	5,100	39,500	110,000	232,000
Thorium, short tons	1,100	8,800	27,000	60,000

Table 5. Known reserves of high-grade uranium and thorium

Country	Mineral	Lb per ton of ore	Short tons
Uranium, U_3O_8			
United States	Carnotite, autunite	5	230,000
Canada	Brannerite	2	400,000
South Africa	Pitchblende (in gold ore)	½	400,000
France			25,000
Other	Davidite, pitchblende		20,000
Total			1,075,000
Thorium, ThO_2			
United States	Thorite, monazite	65	20,000
Canada	Brannerite	1	200,000
India	Monazite	200	150,000
Brazil	Monazite	130	10,000
Other			20,000
Total			400,000

* After concentration

The inventory requirements per Mw(e) are 0 5 tons natural uranium or equivalent (converters); 4 8 kg U^{235} or Pu^{239} plus 0.6 tons U^{238} (fast breeders); 1.1 kg U^{235} or U^{233} plus 0 3 tons thorium (thorium breeders) The fuel consumption or production per MwY(e) is 0 06 tons natural uranium (converters); 0.5 kg Pu^{239} (fast breeders); 0.16 kg U^{233} (thorium breeders)

Reserves of uranium and thorium. Known reserves of high-grade ores from which uranium and thorium can be recovered at $10/lb (of oxide) or less are summarized in Table 5

Estimates of additional reserves of high-grade uranium ores based on general geologic data and the discovery experience since 1950 indicate that the total uranium reserves in non Communist countries may contain as much as 4,000,000 tons of U_3O_8. Lower-grade ores such as bituminous shale and phosphate deposits, wh[illegible] in ⅒ [illegible] of uranium per ton, are

action of the dissolution medium. Except for a few fission products which are volatilized during dissolution, all the constituents initially in the fuel are obtained in the dissolver solution as soluble salts. This solution may be of very complex composition, because in addition to nitric acid, other chemicals may be added to promote dissolution of fuels which resist chemical attack. Following the dissolution step, it is generally necessary to treat the resulting solution by various means in order to accommodate its use as a feed solution to the solvent extraction process. The two most important reasons for such pretreatment are (1) adjustment of oxidation states and concentrations of solution constituents for optimum recovery and purification performance in the solvent extraction process, and (2) modification of corrosion behavior of the solution toward materials employed in process equipment. In some instances, the pretreatment may include a simple type of process step, such as a selective precipitation, to remove the bulk of some specific impurities, for example, certain fission products or dissolved structural material.

Solvent extraction. The next operation is the solvent-extraction step. It is here that actual recovery and purification are performed by the use of special organic solvents. By far the most frequently used solvent is a mixture of tributyl phosphate (a chemical used in the paint industry and usually abbreviated as TBP) and kerosine. The basic principles by which separation is achieved during solvent extraction are immiscibility of the organic solvent with the aqueous solution of irradiated fuel, and differences with which components, initially present in the aqueous fuel solution, distribute or partition themselves between the organic and the aqueous solutions when the organic solution is first thoroughly stirred or mixed with the aqueous solution and is later separated from the aqueous solution.

If a quantity of suitably prepared solution of irradiated fuel is mixed with a similar quantity of TBP-kerosine solution and is allowed to stand, the following results will occur. The TBP-kerosine mixture will locate itself essentially quantitatively above the aqueous solution because the organic mixture is not miscible with the aqueous solution and is less dense than the aqueous solution. If analyses are performed on the separated liquids, it will be found that a very appreciable portion of the uranium and plutonium (and thorium if it is also present) have transferred to the organic mixture, but only a minute fraction of the fission products and other impurities have transferred. In order to enhance the transfer of uranium and plutonium into the organic mixture without influencing the transfer of fission products appreciably, it is customary to have present in the aqueous solution large concentrations of certain chemicals called salting agents. If the organic solvent containing the uranium and plutonium is now brought into contact with an aqueous solution wherein salting agents are absent, the uranium and plutonium will have retransferred almost quantitatively to the new aqueous solution. The solvent can then be reused. Because only a minute fraction of the fission products was initially transferred to the organic mixture, the new aqueous solution contains recovered uranium and plutonium well separated from fission products. It is possible to take this aqueous solution and separate the uranium and plutonium from each other as is sometimes desired. This is also done with the same solvent by taking advantage of the fact that under certain conditions (reduced oxidation state of plutonium) plutonium extraction by the solvent is very small. Thus separation of the two heavy elements is achieved in much the same way that the impurities (for example, fission products) were initially separated from uranium and plutonium.

Batch extraction. If the initial fuel solution is repeatedly treated with quantities of fresh solvent, it is possible, in principle, to recover all of the uranium and plutonium from the fuel solution, leaving behind essentially all of the fission products and other impurities. This type of solvent extraction procedure is called a multiple batch extraction. Although such a procedure is sometimes used for laboratory purposes, in actual practice of fuel reprocessing on a large scale it is more convenient, efficient, and economic to employ a procedure called continuous countercurrent extraction. The basic principles of separation of components with the latter procedure are still the same as with the batch type.

Countercurrent extraction. The importance of continuous countercurrent solvent extraction in fuel reprocessing merits further elaboration. As its name implies, it is a continuous operation conducted to obtain repeated mixing and separation of the organic solvent and the aqueous solution of irradiated fuel from which it is desired to remove all of the valuable products freed of impurities. The continuous nature of operation is also applied in the step wherein the purified products are retransferred to an essentially pure water solution (that is, free of salting agents). Continuous separation of plutonium and uranium (or thorium and uranium in the case of thorium-uranium-fueled reactors) can also be performed. The principal advantages of continuous over batch operations are more uniform product quality, greater ease of instrumentation, and less severe problems in avoiding accidental accumulation of sufficient fissionable material in one location to cause a nuclear reaction. The countercurrent aspects of the operation are derived from having the organic solvent flow in a direction opposite from that of the aqueous solution. This allows maximum loading of the organic mixture with the components to be extracted because fresh solvent encounters initially low concentrations of these components and progressively higher concentrations as the solvent moves toward the point at which the aqueous solution is introduced. In this way, a minimum of solvent is required to achieve maximum recovery of desired

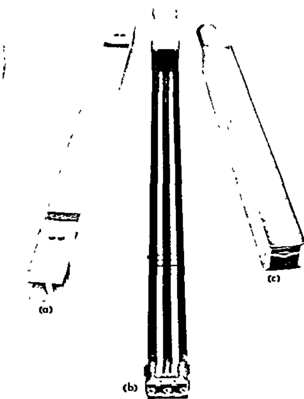


Fig 3. Fuel element. (a) Plate-type. (b) Rod-type (c) Plate-type (*From Robert Laws, Salon of Photography*)

be chemically compatible with the material and absolutely leak-tight to prevent the release of radioactive fission products and chemical reaction of the uranium with coolant. Typical fuel elements include aluminum-clad uranium metal rods for plutonium production reactors; zirconium-clad plates containing U-Zr alloy or stainless-clad UO_2 dispersed in stainless steel for propulsion reactors; and zirconium- or stainless-steel-clad UO_2 pellets for *central station power reactors* *See* FISSION, NUCLEAR; NEUTRON, NUCLEAR POWER; PLUTONIUM; RADIOACTIVITY; REACTOR, NUCLEAR (CLASSIFICATION), THERMONUCLEAR REACTION; THORIUM; URANIUM. [J.A L.]

Bibliography: Atomic Industrial Forum, *Growth Survey of the Atomic Industry, 1958–1968*, 1958;

ress in Nuclear Energy, ser. 3, vol. 1, 1958; W. K. Davis, Forecasts of installed nuclear capacity, *Nucleonics*, 15(4):18, 1957; H. Etherington (ed.), *Nuclear Engineering Handbook*, 1958; S. Glasstone, *Sourcebook on Atomic Energy*, 2d ed., 1958; J. C. Johnson, Resources of nuclear fuel for atomic power, *Proc. Inter. Conf. Peaceful Uses of Atomic Energy*, 2:3–6, 1958; J. A. Lane, Growth potential of U.S. nuclear power industry, *Nucleonics*, 12(6):12-18, 1954; R. Stephenson, *Introduction to Nuclear Engineering*, 2d ed., 1959; W. D. Wilkinson and W. F. Murphy, *Nuclear Reactor Metallurgy*, 1958.

Nuclear fuels reprocessing

The periodic chemical, physical, and metallurgical treatment of materials used as fuel elements in nuclear reactors. In the operation of a nuclear reactor, it is almost invariably necessary to discharge the fuel well before its complete consumption has occurred because of physical or chemical damage incurred by the fuel, loss of capability of residual fuel to maintain the nuclear chain reaction, and desire to recover newly produced fissionable material. *See* NUCLEAR FUELS.

For whichever of these reasons the fuel has been discharged, chemical reprocessing is conducted to recover and purify the residual fissionable constituents. If fertile material is also contained in the fuel, this, too, is ordinarily recovered and purified during *fuel reprocessing. Purification of the* valuable constituents consists of the removal of fission products and extraneous structural material present in the fuel.

Because of the frequency of fuel discharge and because of the extreme value of fissionable materials (for example, about $7000/lb), it is important that the degree of recovery approach 100% as closely as practicable. With regard to purification it is commonly necessary to reduce the fission product impurity content of discharged fuel by a factor of 10^7 in order to make the recovered material safe to handle during refabrication into new fuel for re-use

There are several basic steps involved in fuel reprocessing. After fuel has been discharged from a nuclear reactor, it is common practice to store the *fuel submerged in 15–20 ft of water* (for cooling and radiation-shielding purposes) for a period of 50–150 days to allow the short-lived fission products to decay radioactively. During this period, the radioactivity of the fuel decreases rapidly and substantially, so that when reprocessing is commenced, shielding requirements are *reduced* to practical thicknesses and radiation damage to chemicals or special structural materials in the reprocessing plant can be held to tolerable magnitude Following the cooling period, the fuel is mechanically cut or disassembled into convenient sizes. At this point, the fuel is ready for chemical reprocessing to enable recovery and purification

The specific steps next undertaken depend upon the particular reprocessing method employed to achieve separation of desired products from each other, from fission products, and from extraneous structural materials. Although many separation methods exist, the one which is based on solvent extraction principles is most frequently used for fuel reprocessing Therefore, the discussion of the next sequence of steps will be based on the use of solvent extraction, about which further details will be given later.

Dissolution of spent fuel. The cut-up or disassembled fuel is charged, along with an appropriate aqueous dissolution medium, into a vessel Here the solid fuel is put into solution by

The average of this field lies in the direction of **J** and interacts with the nuclear magnetic moment. The interaction of a *magnetic moment* with a magnetic field is $-\mathbf{\mu} \cdot \mathbf{H}$ and for these purposes can be written as

$$W = -g_I\mu_N \mathbf{I}\cdot\mathbf{H}_e = ha\mathbf{I}\cdot\mathbf{J} \quad (4)$$

Here g_I is the dimensionless g factor for the nucleus and is defined by the relations

$$\mu = g_I I\mu_N \qquad \mathbf{\mu} = g_I\mu_N\mathbf{I} \quad (5)$$

and $\mathbf{H}_e$ is the electronic magnetic field which is parallel to **J**. For one nuclear magneton (one nuclear magneton is equal to $eh/2Mc$, where M is the proton mass, and has the value 5.0504×10^{-24} erg gauss^{-1}), *the constant* a *is of the order of* 1000 Mc/sec for the orders of magnitude assigned (*see* MAGNETON). The total angular momentum is now $\mathbf{F} = \mathbf{I} + \mathbf{J}$, and **F** is also quantized. Since the quantity

$$\mathbf{I} \cdot \mathbf{J} = [\tfrac{1}{2}F(F+1) - I(I+1) - J(J+1)]$$

the energy levels are

$$W_F = \frac{ha}{2}\,[F(F+1) - I(I+1) - J(J+1)] \quad (6)$$

where $F = I + J, I + J - 1, \ldots, I - J$. There are $2I + 1$ or $2J + 1$ F levels, whichever number is smaller. In most experiments, the selection rules for observing differences in these energies are $\Delta F = 0, \pm 1$, and the corresponding frequencies are

$$f_F = \frac{W_F - W_{F-1}}{h} = aF \quad (7)$$

See SELECTION RULES (PHYSICS)

Resonance lines that *obey* this *simple* scheme are said to follow the interval rule. In general, the hyperfine energy can be expressed as a sum of polynomials in $(\mathbf{I} \cdot \mathbf{J})$ of degree n, $n = 1, 2, \ldots$. If $n \leq 1$, the interaction depends only on $\mathbf{\mu}$. If terms like $(\mathbf{I} \cdot \mathbf{J})^2$ are needed to fit the data, they are usually caused by a nuclear quadrupole moment interacting with the gradient of the electric field at the nucleus. The presence of such terms is indicated by a violation of the interval rule. The octupole term is identified by the coefficient of the $(\mathbf{I} \cdot \mathbf{J})^3$ term. This term is very small and may be largely spurious because of electronic perturbations. The highest moment effective in the hyperfine interaction is determined by the smaller of the two integers $2I$ or $2J$. The quadrupole interaction involves the product of the nuclear quadrupole moment Q and the spatial derivative of the electric field at the nucleus due to the other charges. The total interaction is of the order of magnitude of $2e^2Q/r^3$. The quadrupole moment Q is of the order of 10^{-24} cm^2, r can be chosen as a_0, the Bohr radius, and division by h gives 600 Mc/sec as the order of magnitude of the interaction. The complete expression for the quadrupole interaction term in an atom is

$$W_Q = \frac{e^2Qq}{2I(2I-1)J(2J-1)}\,[3(\mathbf{I} \cdot \mathbf{J})^2 + \tfrac{3}{2}(\mathbf{I} \cdot \mathbf{J}) - I(I+1)J(J+1)] \quad (8)$$

Here q is the average of the quantity

$$-\sum_i (3\cos^2\theta_i - 1)r_i^{-3}$$

over the electronic state characterized by the maximum value of J_z, $J_z = J$, and the sum is over each of the electrons, where r_i is the radius vector from the nucleus to the ith electron. The differences between these terms are small compared with the electronic term differences alone, but they *are detected in experiments designed to observe the* change in relative orientation of **I** and **J** in going from one F state to another. The principal methods for studying nuclear moments in free atoms are the resonance atomic beam method and high resolution optical spectroscopy. If $J = 0$, the nuclear magnetic moment is studied by methods similar to those for molecules.

Moments in free molecules. In free molecules, the nuclear moments and spins manifest themselves through hyperfine couplings similar in origin to those of *free atoms. The charge distributions* in *molecules* vary widely, depending upon the nature of the bonding, and therefore the gradient of the field that couples to the quadrupole moment is much more variable, and so the effect is more variable than in the case of atoms. The *magnetic moment terms* also vary *greatly in size, depending upon whether or not the* valence electrons pair, as, for example, in the case of $^1\Sigma$ or $^2\Sigma$ molecules, respectively. If the electrons have an appreciable fraction of their total angular momentum unquenched, the situation is, in order of magnitude, [illegible] terms become very small (less than a fraction of 1 Mc/sec) and the quadrupole terms *dominate.*

The principal methods employed for the study of *nuclear moments in free* molecules are nuclear magnetic resonance, molecular beams, and microwave absorption in gases. The third method is perhaps the most precise for the quadrupole *moments.*

The nuclear moments and angular *momenta* play a very special role in *molecules* that have two or more identical nuclei. Nuclei of even mass number A obey Bose-Einstein statistics; nuclei of odd A obey Fermi-Dirac statistics. The effect of the nuclear statistics in elimination of certain of the possible states is very striking. It shows *experimentally* in the alternating intensity of the optical spectroscopic bands and in altered intensities in microwave spectroscopy. The spins of many nuclei have been determined *this* way. It is particularly powerful for the case $I = 0$, for which all other methods fail in an absolute sense.

Moments in crystals. In crystals *that* have ionic species in very dilute concentration, the paramagnetic resonance absorption at low temperatures yields a structure composed of relatively narrow lines and is *relatively* easy to observe. The situation in order of magnitude is very much like that for the

materials with solvent-extraction equipment of a given efficiency.

With proper process conditions and suitable equipment, the solvent-extraction operation yields nearly complete recovery and purification of products in the form of dilute aqueous solutions. These solutions are subsequently further processed to give finally the form of plutonium, uranium, or thorium which is suitable for reuse in nuclear reactors. In the case of uranium which has been depleted in its U^{235} content, the processing may include isotope reenrichment in gaseous diffusion plants. The fission products and other impurities initially present in the fuel are also obtained in the form of aqueous solution waste. The waste is concentrated by evaporation and then usually neutralized before it is introduced into underground tanks for indefinite storage. *See* SOLVENT EXTRACTION.

Processing plants. The plants in which fuel reprocessing is performed are large and expensive. Their size may range up to a few hundred yards in length and their cost of construction may exceed $20,000,000. The basic reason for this high cost is the enormous thickness of shielding (up to 7 ft of high-density concrete) required to protect the operating personnel from radiation. Operation of the plants, including sampling for process control, is conducted by remote means. In some plants, even the repair and modification of equipment in high-radiation zones are performed by remote techniques. The additional cost of this type of maintenance is large. For those plants in which maintenance is performed by direct methods, the initial capital cost is reduced, but this may be offset to a large extent by increased operating costs when decontamination is difficult and permissible working time of maintenance personnel is limited. Because of the difficulty and cost of maintenance by either remote or direct methods, more spare equipment and higher standards of design, construction, and installation are necessary in fuel-reprocessing plants than in conventional chemical plants. Special precautions which also contribute to increased capital and operating costs must be made in fuel-reprocessing plants to avoid nuclear accidents from inadvertent accumulation of fissionable materials. This is particularly important when highly enriched fuels are reprocessed.

Thus the gross capital and operating costs are high for a fuel-reprocessing plant. The unit cost of recovered products is also very large because the output of moderately large plants is relatively small, being only a few tons per day for very slightly enriched (3% or less) fuel to as little as 10-20 lb/day for highly enriched (about 90%) fuel. Unit cost can be substantially reduced, however, by increased capacity, because total capital and operating costs do not increase proportionally.

Further improvements are being made in solvent-extraction processes to accommodate new fuels and to reduce costs. In addition, other processes are being developed which offer promise of certain advantages over solvent extraction, for example, less susceptibility to radiation problems with fuels that have been cooled for only a short time, and fewer operating steps required.

The operating experience to date with fuel reprocessing plants has shown them to be relatively safe in spite of hazards from radiation and nuclear criticality, and other hazards of more conventional nature. *See* RADIOCHEMICAL LABORATORY. [S.L.]

Bibliography: M. Benedict and T. H. Pigford, *Nuclear Chemical Engineering*, 1957; F. R. Bruce, J. M. Fletcher, H. H. Hyman, and J. J. Katz, *Process Chemistry*, vol. 1, 1956; H. Etherington (ed.), *Nuclear Engineering Handbook*, 1958; *Symposium on the Reprocessing of Irradiated Fuels*, U.S. Comm. TID-7534, 3 vols., 1957.

Nuclear moments

The various static electric and magnetic moments possessed by the atomic nucleus. The nuclear moments that play an important role in research are, in order of complexity, the magnetic dipole moment, the electric quadrupole moment, and the magnetic octupole moment. These are defined respectively by

$$\mu = \int \mathbf{M}_z \, d\tau \tag{1}$$

$$Q = \frac{1}{e} \int (3 \cos^2 \theta - 1) r^2 \rho \, d\tau \tag{2}$$

$$M_3 = \int \frac{5 \cos^3 \theta - 3 \cos \theta}{2} \operatorname{div} \mathbf{M} r^3 \, d\tau \tag{3}$$

The nuclear spin is $\hbar\mathbf{I}$, and its projection on the z axis, $\hbar I_z$, is quantized to have the values $\hbar I$, $\hbar(I-1)$, . . . , $-\hbar I$, where I is a positive integer or half-integer, and $\hbar$ is Planck's constant h divided by 2π. The integrals of Eqs. (1), (2), and (3) are to be taken so that the physical quantities involved are for the nuclear state $\hbar I_z = \hbar I$, which is the largest positive value of the projection of the nuclear angular momentum on the z axis. This is the position most closely identified with the classical spin parallel to the z axis. *See* SPIN (QUANTUM MECHANICS).

The quantity r is the radius vector from the center of the nucleus to the element of nuclear volume $d\tau$; ρ is the charge density in the nucleus; $\mathbf{M}$ is the magnetization in the nucleus; θ is the angle between r and the z axis; and e is the electronic charge.

Moments in free atoms. Nuclear spins and moments manifest themselves in free atoms through the hyperfine interaction, that is, a close splitting of some of the energy levels of the atom with resulting hyperfine splitting of the spectral lines (*see* HYPERFINE STRUCTURE). If the nuclear spin is zero, the angular momentum of the atom is solely the result of the electronic structure, and the geometric aspects of the electric and magnetic fields of the electrons as well as their magnitudes, are determined by the total angular momentum $\mathbf{J}$. (All angular momenta will from now on be expressed in units of $\hbar$.) For example, if $\mathbf{J} \neq 0$, the magnetic field at the nucleus due to the electrons lies betw[illegible] 10^5 and [illegible]

ences as a result of quadrupole interactions alone in zero magnetic field. In that case, the term pure quadrupole resonance is used.

Paramagnetic resonance This is observed in crystals by placing the crystal in a cavity, which is part of a resonant circuit, and measuring the power absorption as an external field is varied through the resonances while the frequency is held constant Since, basically, a magnetic moment of the order of a Bohr magneton is involved in the transition, the method is far more sensitive in terms of number of atoms than is the nuclear resonance method. The experiments must be conducted in cryostats operating at temperatures in the range of that of liquid helium This is primarily necessary to give relaxation times that are long enough to result in narrow lines for reasonable precision A secondary purpose is to supply a large difference in population between levels and thus enhance the signal. This temperature effect on the signal is, of course, common to all methods that detect by power absorption

Microwave spectroscopy. In gases, microwave spectroscopy is performed by observation of the attenuation of a signal propagated in a wave guide containing the absorber at low pressures The observations are usually the electric dipole transitions between the different rotational states of the molecules These states are about 20 000 Mc/sec apart, and each level is split by quadrupole hyperfine interactions of the order of 100 Mc/sec As a result a complex pattern of lines will be observed as the frequency is varied The lines are quite sharp and very precise quadrupole couplings are obtained, as well as precise information on rotational and vibrational constants of the molecule as a whole, and also on nuclear spin

Optical spectroscopy In the visible and ultraviolet, optical spectroscopy continues to be a prolific source of information on the spins and hyperfine couplings of elements The effect is the splitting of spectral terms by the hyperfine interactions of Eqs (6) and (8). It has special value because of the variety of states other than the atom ... pr th re ... poor resolution and the effect of the isotope shift This is an absolute shift in levels between different isotopes because of the slightly different nuclear radii and charge distributions (see ISOTOPE SHIFT) Optical spectroscopy has a special advantage over other methods in that the intensity pattern of the lines often yields the sign of the interaction constant *a* See ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY; MAGNETIC RELAXATION, MAGNETIC RESONANCE, MICROWAVE SPECTROSCOPY, MOLECULAR BEAMS; MOLECULAR STRUCTURE AND SPECTRA; NUCLEAR STRUCTURE

[W.A N.]

Bibliography. S Fluegge (ed), *Handbuch der Physik*, vols. 37–38, 1958–1959, H Kopfermann, *Nuclear Moments*, 1958; W. A Nierenberg, The measurement of the nuclear spins and static moments of radioactive isotopes, *Ann. Rev. Nuclear Sci.*, 7:349–406, 1957; N. F. Ramsey, *Molecular Beams*, 1956; K. F. Smith, Nuclear moments and spins, *Progr in Nuclear Phys*, 6:52–107, 1957.

Nuclear physics

Nuclear physics can be divided into two classifications, low-energy and high-energy. The first variety is concerned primarily with the arrangement of the protons and neutrons within the atomic nucleus, and with the nature of the forces between these nuclear particles. Low-energy means several million electron volts (Mev), because the excited states of nuclei range up to 20 Mev, and the binding energy per nuclear particle is about 8 Mev. In high-energy nuclear physics, on the other hand, particle energies of hundreds of Mev or even several billion electron volts (Bev) are used to produce mesons and so-called strange particles The interactions of these elementary particles are then studied Cosmic rays fall into the high energy classification, because the primary cosmic rays are largely protons with energies of many Bev.

The rapid growth of nuclear physics has come about because of the possibility of tremendous technical exploitation. Experimentalists have built a great variety of particle detectors, uranium-fueled reactors, and particle accelerators ranging up to the huge 25-Bev proton synchrotrons. Theorists have as their goals improvements in the knowledge of nuclear forces, a comprehensive theory of nuclear structure, and a satisfactory generalized field theory. See ELEMENTARY PARTICLE and the articles listed therein; see also COSMIC RAYS; FISSION, NUCLEAR; FUSION, NUCLEAR; ISOTOPE ... DETECTOR; RADIOACTIVITY, REACTOR, NUCLEAR, SCATTERING EXPERIMENTS, NUCLEAR.

[W.W W.]

Nuclear power

Power (or energy) derived from the fusion of the nuclei of low-atomic-weight elements, such as deuterium or tritium, or the fission (splitting) of the nuclei of heavy elements, such as uranium. The amount of energy per atom released in fission and fission reactions exceeds the amount for chemical reactions such as the combustion of coal by factors of several millions. In the case of fission, high energy input is required to initiate and sustain the reaction See ENERGY SOURCES; FISSION, NUCLEAR; FUSION, NUCLEAR.

Advantages of nuclear power. The primary advantage of nuclear power is that nuclear fuels are more abundant than fossil fuels and comprise a more concentrated source of energy The fission of 1 lb of nuclear fuel, for example, liberates an amount of energy equivalent to the ...

free atom. The dilution referred to means that the ions are sufficiently removed from one another that they no longer effectively interact directly with one another. They are in homologous sites in the crystal lattice, so they behave alike in the power absorption spectrum. The situation is more complicated in that the crystalline electric field, as a whole, is usually strong enough to compete with the spin-orbit coupling of the electrons in the ion, or even stronger, so that the energy-level classification of the ion is more complicated than if it were free. Nevertheless, there are often well-separated electronic ground states that have a remaining degeneracy (usually two-fold) that corresponds to the $2J + 1$ degenerate states of a free ion. This degeneracy can be partially removed by the hyperfine splitting caused by the interaction of the nuclear moments with the fields due to the ion A difference between this and the atomic case is that the nucleus can also interact with the crystalline electric fields via its quadrupole moment This method of paramagnetic resonance in crystals has been very important for the measurement of spins and moments of rare species because of its great sensitivity.

As in the free atom, the coefficients of μ and Q in the interaction energies that are observed involve averages of r^{-3} over the ion with respect to the electron distribution This important quantity in the case of the atom, molecule, or crystal yields valuable information on the nature of the bonding There are gases such as N_2 and O_2 which are $^3\Sigma$ in their ground states. The electronic angular momentum couples to the nuclear moments to yield a hyperfine structure, which has also been observed by paramagnetic resonance.

Measurement of nuclear moments. The measurement of nuclear moments can be made in the physical situations described in the foregoing paragraphs by a wide variety of methods An external magnetic field can be applied that interacts directly with the electronic moment, if it exists, and with the nuclear moment, and this removes the $2F + 1$ degeneracy that remains in each of the hyperfine structure terms. In paramagnetic resonance in crystals and gases, in atomic beams, and in related methods, the magnetic moments are most often inferred from ratios of the magnetic hyperfine constant a with respect to some known and calibrated isotope. The direct interaction of the nuclear moment with the external field can be observed, but it is a relatively small term and the accuracy may be limited. If no isotope of the same material has a known nuclear g factor and a known hyperfine constant a, the direct nuclear interaction must be observed or the atomic fields must be calculated The latter procedure is fraught with considerable uncertainty. Microwave electric dipole absorption in gaseous molecules yields excellent values of quadrupole interactions, as do paramagnetic resonance in crystals and atomic beams Other methods are molecular beams using magnetic resonance and magnetic focusing or electric resonance and electric focusing. There is no direct method for observing quadrupole moments, and so the couplings with internal atomic, molecular, and crystalline fields must be interpreted. In general the results are somewhat uncertain except for the case of the deuteron.

Nuclear resonance; molecular beams. The nuclear magnetic moment can be measured by nuclear magnetic resonance in solids, liquids, and gases, and by molecular beams. If the situation in each case is that the internal magnetic fields and derivatives of the electric fields of the nearby electrons are quenched because of pairing of the electrons, the interaction of the nuclear magnetic moment with the external field is dominant Since the nucleus has an angular momentum and suffers a torque because of the effect of the magnetic field on the moment, it will precess about the field at a constant frequency. If an rf magnetic field is applied at this frequency, power will be absorbed by this precessing magnet and it will change its inclination. The interaction energy is, from Eq (4),

$$W = -g_I\mu_N H I_z = -g_I\mu_N H m_I \qquad (9)$$

where m_I is the quantum number for I_z, and since the rule of combination is $\Delta m_I = \pm 1$, the frequency for transitions is

$$f = g_I\mu_N H/h \qquad (10)$$

The molecular beam apparatus is designed to detect the change in orientation, while the nuclear magnetic resonance apparatus is a bridge circuit designed to detect the absorbed power. The frequencies for fields of 5000 gauss are about 5 Mc/sec The field can be calibrated in a variety of ways, but it has become standard to use the resonance of the proton. Very careful absolute determinations of the g factor for the proton have been made, and great accuracy is attainable by this method Sensitivities of 1 part in 10^8 are possible under optimum circumstances The atomic beam method is a variant in that essentially the unpaired electron is flipped, and its change in orientation is detected by noting the change of trajectory of the atom in an inhomogeneous magnetic field. This is similar to the nuclear reorientation detection of the molecular beam apparatus. By virtue of hyperfine interactions such as those of Eqs (7) and (8), the frequencies for these transitions are modified, and the constants a and b are measured. For Cs^{133}, the constant a has been measured to parts in 10^{10} ($a = 2{,}298{,}157{,}913 \pm 5$ cps) and is being used as a time standard. See ATOMIC CLOCK; ELECTRICAL STANDARDS.

Both the molecular beam technique and the nuclear resonance technique are used to measure quadrupole interactions, when these are present in sufficient magnitude. In both cases, the quadrupole interactions are represented by a broadening of the resonance line above the natural widths and a definite structure which is characterized by the specific value of the nuclear s in It is possil o observe a resonance assoc ener

ences as a result of quadrupole interactions alone in zero magnetic field. In that case, the term pure quadrupole resonance is used.

Paramagnetic resonance. This is observed in crystals by placing the crystal in a cavity, which is part of a resonant circuit, and measuring the power absorption as an external field is varied through the resonances while the frequency is held constant. Since, basically, a magnetic moment of the order of a Bohr magneton is involved in the transition, the method is far more sensitive in terms of number of atoms than is the nuclear resonance method. The experiments must be conducted in cryostats operating at temperatures in the range of that of liquid helium. This is primarily necessary to give relaxation times that are long enough to result in narrow lines for reasonable precision. A secondary purpose is to supply a large difference in population between levels and thus enhance the signal. This temperature effect on the signal is, of course, common to all methods that detect by power absorption.

Microwave spectroscopy. In gases, microwave spectroscopy is performed by observation of the attenuation of a signal propagated in a wave guide containing the absorber at low pressures. The observations are usually the electric dipole transitions between the different rotational states of the molecules. These states are about 20 000 Mc/sec apart, and each level is split by quadrupole hyperfine interactions of the order of 100 Mc/sec. As a result, a complex pattern of lines will be observed as the frequency is varied. The lines are quite sharp, and very precise quadrupole couplings are obtained, as well as precise information on rotational and vibrational constants of the molecule as a whole and also on nuclear spin.

Optical spectroscopy. In the visible and ultraviolet, optical spectroscopy continues to be a prolific source of information on the spins and hyperfine couplings of elements. The effect is the splitting of spectral terms by the hyperfine interactions of Eqs. (6) and (8). It has special value because of the variety of states other than the atomic [illegible]
poss[illegible]
the [illegible]
rela[illegible] poor resolution and the effect of the isotope shift. This is an absolute shift in levels between different isotopes because of the slightly different nuclear radii and charge distributions (see ISOTOPE SHIFT). Optical spectroscopy has a special advantage over other methods in that the intensity pattern of the lines often yields the sign of the interaction constant *a*. See ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY; MAGNETIC RELAXATION; MAGNETIC RESONANCE; MICROWAVE SPECTROSCOPY; MOLECULAR BEAMS; MOLECULAR STRUCTURE AND SPECTRA; NUCLEAR STRUCTURE.

[W.A.N.]

Bibliography: S. Fluegge (ed.), *Handbuch der Physik*, vols. 37–38, 1958–1959; H. Kopfermann, *Nuclear Moments*, 1958; W. A. Nierenberg, The measurement of the nuclear spins and static moments of radioactive isotopes, *Ann. Rev. Nuclear Sci.*, 7:349–406, 1957; N. F. Ramsey, *Molecular Beams*, 1956; K. F. Smith, Nuclear moments and spins, *Progr. in Nuclear Phys.*, 6:52–107, 1957.

Nuclear physics

Nuclear physics can be divided into two classifications, low-energy and high-energy. The first variety is concerned primarily with the arrangement of the protons and neutrons within the atomic nucleus, and with the nature of the forces between these nuclear particles. Low-energy means several million electron volts (Mev), because the excited states of nuclei range up to 20 Mev, and the binding energy per nuclear particle is about 8 Mev. In high-energy nuclear physics, on the other hand, particle energies of hundreds of Mev or even several billion electron volts (Bev) are used to produce mesons and so-called strange particles. The interactions of these elementary particles are then studied. Cosmic rays fall into the high energy classification, because the primary cosmic rays are largely protons with energies of many Bev.

The rapid growth of nuclear physics has come about because of the possibility of tremendous technical exploitation. Experimentalists have built a great variety of particle detectors, uranium-fueled reactors, and particle accelerators ranging up to the huge 25-Bev proton synchrotrons. Theorists have as their goals improvements in the knowledge of nuclear forces, a comprehensive theory of nuclear structure, and a satisfactory generalized field theory. See ELEMENTARY PARTICLE and the articles listed therein; see also COSMIC RAYS; FISSION, NUCLEAR; FUSION, NUCLEAR; ISOTOPE; NUCLEAR CHEMISTRY; NUCLEAR MOMENTS; NUCLEAR REACTION; NUCLEAR STRUCTURE; PARTICLE ACCELERATOR; PARTICLE DETECTOR; RADIOACTIVITY; REACTOR, NUCLEAR; SCATTERING EXPERIMENTS, NUCLEAR.

[W.W.W.]

Nuclear power

Power (or energy) derived from the fusion of the nuclei of low-atomic-weight elements, such as deuterium or tritium, or the fission (splitting) of the nuclei of heavy elements, such as uranium. The amount of energy per atom released in fission and fission reactions exceeds the amount for chemical reactions such as the combustion of coal by factors of several millions. In the case of fission, high energy input is required to initiate and sustain the reaction. See ENERGY SOURCES; FISSION, NUCLEAR; FUSION, NUCLEAR.

Advantages of nuclear power. The primary advantage of nuclear power is that nuclear fuels are more abundant than fossil fuels and comprise a more concentrated source of energy. The fission of 1 lb of nuclear fuel, for example, liberates an amount of energy equivalent to the [illegible]

nuclear power is particularly attractive for locations where fuel transportation costs are high. This includes ship propulsion and power production at locations far removed from supplies of conventional fuels. Moreover, because of their high energy content, nuclear fuels offer the promise of eventually producing power and electricity at costs less than that from fossil fuels.

Estimates of the world reserves of uranium and thorium recoverable at present-day prices ($10/lb) indicate an energy reserve about ten times that of economically recoverable fossil fuels (*see* NUCLEAR FUELS). When these supplies are exhausted, the vastly greater supply of uranium from lower-grade sources can be utilized or, as an alternative, deuterium can be recovered from water and burned in a thermonuclear (fusion) reactor. Because sea water contains some 10^{17} lb of deuterium, this latter source of energy is almost inexhaustible.

Disadvantages of nuclear power. The extraction of energy from nuclear fuels is considerably more complicated than the combustion of fossil fuels. Before uranium can be used in a nuclear reactor, it must be recovered from ore (which normally contains only ½–5 lb of uranium oxide per ton), purified, and fabricated into some sort of fuel element or fuel material. In addition, it is not possible to achieve 100% fission of the uranium because of losses during processing, holdup in the reactor and processing plants, and nonfission nuclear reactions. Depending upon the type of reactor and its associated processing cycle, the utilization of uranium may be limited to 2–30% of the total amount mined.

In the case of the fusion fuels, 1 gal of ordinary water contains only 0.125 g of deuterium, which must be extracted before it can be used. However, the cost of producing this amount of deuterium, which has an energy equivalent to more than 200 gal of gasoline, amounts to only about 4¢ at the 1958 price of heavy water, $28/lb.

Power generation. The heat generated in the nuclear fuel due to the fission process and subsequently transferred to a coolant can be recovered by using the coolant either to produce steam (indirect cycle) or as the working fluid for running a turbomotor (direct cycle). These cycles are depicted schematically in Fig. 1. The indirect cycle has the advantage of being applicable to a wide variety of coolants and coolant temperatures and this method eliminates the need for shielding the power-generating equipment for protection from radioactive products in the coolant. The direct cycle has the advantage of a higher thermal efficiency for a given pressure. The development of cladding materials capable of handling high-temperature coolants, such as gases or superheated steam, however, increases the attractiveness of the direct-cycle approach, and a number of nuclear power applications using such a direct cycle are being actively pursued.

For the most efficient utilization of the reactor-produced heat, coolant temperatures should be as high as possible, subject only to the physical

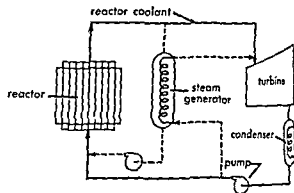


Fig. 1. Power cycles for nuclear plants.

limitations of the materials in the reactor. Unlike fossil-fuel-burning plants, where temperatures are limited to those of the flame, nuclear fuel temperatures can rise practically without limit. Although this means that loss of coolant can result in excessive fuel temperatures, it also means that there is a possibility of developing a nuclear-power boiler with a considerably higher thermal efficiency than a coal-fired boiler. The ability to achieve high temperature operation is a prime requisite of reactors used for the production of electricity or mobile power. This is in contrast to research reactors and production reactors, where low temperature operation is usually desirable.

Central-station electrical plants. For central-station power production, a number of factors other than coolant temperatures affect the cost of power and must, therefore, be considered in the design of such plants. The most important of these factors are (1) the neutron economy of the fuel system, which governs the lifetime of fuel elements and the amount of by-product fissionable material produced, and (2) the amount of power that can be extracted from a reactor of given size and given amount of fuel and fertile material, which influences the capital cost of the plant.

Types of solid-fuel reactors that most satisfactorily combine these factors are (1) H_2O moderated reactors cooled with pressurized H_2O or boiling water (steam), (2) D_2O moderated reactors cooled with D_2O (heavy water), helium, CO_2, organic coolant (terphenyl), or sodium; (3) graphite-moderated reactors cooled with pressurized H_2O, helium, CO_2, or sodium, and (4) unmoderated, sodium-cooled fast reactors.

Types of fluid fuel reactors under development include (1) D_2O-moderated homogeneous systems, (2) graphite-moderated liquid bismuth-uranium fueled; and (3) graphite moderated or homogeneous molten-fluoride systems.

Economics of central-station power. The cost of producing electricity in a nuclear power plant depends on the capital cost of the plant, the annual fixed charges on this capital cost (that is, return on investment, taxes, insurance, and depreciation); fuel-cycle costs; and operating, labor, and maintenance costs.

Costs of a nuclear rea[illegible] [illegible]gned t[illegible] steam for power pro[illegible]

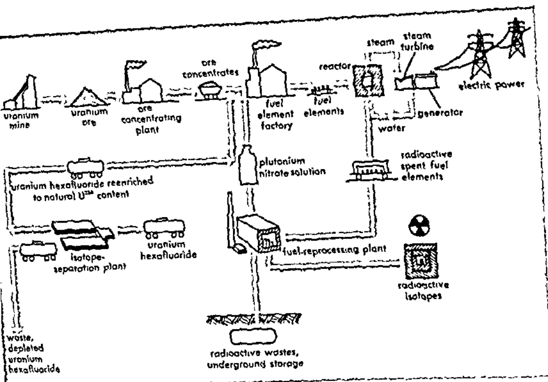


Fig 2 Operations associated with nuclear power plant (From M Benedict and T Pigford, Nuclear Chemical Engineering, McGraw-Hill, 1957)

size, and coolant performance Unit costs of small power reactors and their associated facilities, for plants in the capacity range of 10,000–20,000 kilowatts (kw) of electricity, lie in the range of \$600–\$800/kw, if built according to present-day technology On the same basis, minimum unit costs of large-scale reactors, 100–200 megawatts of electricity, are in the range of \$150–\$200/kw Corresponding costs of the steam-producing portion of a modern coal- or oil-fired plant are in the range of \$50–\$60/kw

The costs for the turbine, generator, and miscellaneous electrical equipment, which are about the same for nuclear and conventional plants, vary from \$110 to \$170/kw, depending on the turbine-plant size and steam temperatures

The costs associated with the nuclear fuel cycle depend on the total number of kilowatt-hours of electricity that are produced from a given fuel element before it has to be replaced and on the net value of recoverable fissionable material in the spent fuel element Fuel costs in a nuclear plant, therefore, are a function of the cost of fissionable material in fresh fuel minus its value in spent fuel, the cost of making fuel elements and reprocessing them after burn-up; and miscellaneous costs due to fuel losses during processing, storage, and shipping of new and spent fuel elements. The charges due to the inventory of fuel, fertile material, and D_2O in the reactor and processing plants are at present fixed by the Atomic Energy Commission at 4% of their value per year

Nuclear power costs. Estimated nuclear-power costs in typical large-scale nuclear plants starting operations in 1963–1965 are summarized in the table

It is evident from the table that, on the basis of estimated power costs, no one reactor type shows at present a decided advantage Advances in technology are predicted for all of these types, moreover, which will reduce power costs to 8 mills/kwhr or less by 1970

Power plants for propulsion. The design of reactors for the propulsion of naval vessels and air-

Estimated nuclear power costs in 100–200 Mw(e) nuclear plants in the United States

Cost	Reactor type		
	Pressurized water	Boiling water	Gas-cooled
Capital cost, \$/kw			
Reactor	205	165	200
Turbogenerator plant	125	125	170
Fixed charges, mills/kwhr			
At 14% per year, 80% load factor	6.6	5.8	7.4
Fuel cycle costs, mills/kwhr			
Fuel inventory	1.3	0.8	1.2
Fabrication	1.1	1.7	0.5
Reprocessing	0.4	0.3	0.3
Uranium burn-up less plutonium credit	1.2	1.0	0.8
Total fuel cost mills/kwhr	4.0	3.8	2.8
Labor and maintenance, mills/kwhr	1.0	1.0	1.0
Total power cost, mills/kwhr	11.0	10.6	11.2

craft involves a consideration of physical size, which in many cases is more important than the economic factors associated with stationary plants. This is because the weight of the shield increases with the diameter of the reactor core. Because even the most compact cores require 50–100 tons of shielding, the problem of transporting this heavy shield is serious. Nuclear systems with as high a power density (power output per unit volume) as possible are, therefore, desired for mobile power reactors. Nuclear propulsion offers the advantages of lower-weight fuel, and less frequent fueling. *See* NUCLEAR AIRCRAFT PROPULSION; NUCLEAR ROCKET; REACTOR, SHIP PROPULSION

Miscellaneous reactor applications. The heat and radiations from the fission reaction can be utilized in a number of ways other than for electricity generation and propulsion. Reactor-produced steam can be used for chemical processes or for heating buildings. Reactor heat and radiation are capable of promoting chemical reactions Radioactive gamma emitters produced in the reactor can sterilize food and other products. Although such applications are all technically practicable, their extensive use will depend on costs of nuclear compared to conventional processes For a discussion of reactor types, *see* REACTOR, NUCLEAR (CLASSIFICATION). [J.A.L.]

Bibliography: C. F. Bonilla (ed.), *Nuclear Engineering*, 1957; H Etherington (ed), *Nuclear Engineering Handbook*, 1958; S. Glasstone, *Principles of Nuclear Reactor Engineering*, 1955; D. B Hoisington, *Nucleonics Fundamentals*, 1959; R. L Murray, *Introduction to Nuclear Engineering*, 1954, R. Stephenson, *Introduction to Nuclear Engineering*, 2d ed., 1958.

Nuclear radiation

A term used to denote all the particles and radiations which emanate from the atomic nucleus as a result of radioactive decay and nuclear reactions. The term was originally used to denote only the ionizing radiations observed from naturally radioactive materials These were α-rays (high-speed helium nuclei), β-rays (negative electrons or negatrons), and γ-rays (electromagnetic radiation of much shorter wavelength than visible light).

The distinction between nuclear radiations and others with similar physical properties lies in whether a nuclear process is involved in their production. Thus, although γ-rays and x-rays are both electromagnetic radiations and, for the same wavelength, are not distinguishable physically, one (γ-radiation) is emitted as a result of a rearrangement of protons and neutrons within the nucleus, while the other (x-rays) results from rearrangement of electrons outside the nucleus.

In addition to α-, β-, and γ-rays, other commonly encountered nuclear radiations are positively charged electrons (positrons), protons, and neutrons. Another radioactive decay product is the neutrino; it is not ordinarily considered as nuclear radiation, since its interaction with matter is very slight. Nuclear reactors are excellent sources of neutrinos. *See* NEUTRINO; NUCLEAR REACTION.

[W.W.E]

Nuclear radiation (biology)

Nuclear radiations are used in biology because of their common property of ionizing matter. This makes their detection relatively simple, or makes possible the production of biological effects in a living cell. Nuclear radiations originate in atomic nuclei, either spontaneously, as in radioactive substances, or through interactions with neutrons, photons, and so on. Gamma radiation originates in atomic nuclei and constitutes one kind of nuclear radiation, but it is otherwise indistinguishable in its effects from x-radiation produced by extranuclear reactions. Because x-rays have been readily available for many years, they have been used more extensively in biology and medicine than γ-rays. Therefore, x-rays must be included in a discussion of the biological and medical uses of nuclear radiations. *See* GAMMA RAYS; NEUTRON; PHOTON; X-RAY(S), PHYSICAL NATURE OF.

Ionizing radiation. Ionizing radiation is any electromagnetic or particulate radiation capable of producing ions, directly or indirectly, in its passage through matter.

Electromagnetic radiations X-rays and γ-rays are electromagnetic radiations, traveling at the speed of light as packages of energy called photons They ionize indirectly by first ejecting electrons at high speed from the atoms with which they interact, these secondary electrons then produce most of the ionization associated with the primary radiation.

Particulate radiation. Fast neutrons are particulate radiation consisting of nuclear particles of mass number 1 and zero charge, traveling at high speed. They ionize indirectly, largely by setting in motion charged particles (from the atomic nuclei) with which they collide Slow or thermal neutrons ionize indirectly by interacting with nuclei, in a process known as neutron capture, to produce ionizing radiation.

Alpha rays are particulate radiation consisting of helium nuclei traveling at high speed Since they are charged particles, they ionize directly Alpha particles are emitted spontaneously by some radioactive nuclides or may result from neutron capture; for example, neutron capture by boron-10 produces lithium-7 and an α-particle The energy of α-particles emitted by radioactive substances is of the order of a few million electron volts (Mev), but α-particles of very much higher energy may be produced in cyclotrons or other particle accelerators from helium-ion beams With such machines other ionizing particles of very high energy, such as protons, deuterons, and so on, may also be produced *See* ALPHA RAYS

Beta rays are particulate radiation consisting of electrons or positrons emitted from a nucleus during β-decay and traveling at high speed. Since they are charged particles, — or

ionize directly. Electron beams of very high energy may be produced by high voltage accelerators, but in that case they are not called β-rays. A pair consisting of one electron and one positron may be formed by one high-energy (1022-mev) photon when it traverses a strong electric field, such as that surrounding a nucleus or an electron. Subsequently, the positron and another electron react, and their mass is transformed into energy in the form of two photons traveling in opposite directions This is called the annihilation process and is the inverse of the pair production process mentioned previously. Ionizing radiations, such as protons, deuterons, α-particles, and neutrons, may be oduced simultaneously by spallation when a very gh-energy particle collides with an atom. In a otographic emulsion or in a cloud chamber, the nizing particles originating from a common point rm stars. *See* BETA RAYS.

Fission occurs in certain heavy nuclei spontaneusly or through interaction with neutrons, charged articles, or photons, and it results in the division f the nucleus into two approximately equal parts. hese fission fragments are endowed with very arge amounts of kinetic energy, carry large posiive charges, and produce very dense ionization in heir short passage through matter

Primary cosmic rays probably consist of atomic nuclei, mainly protons with extremely high energies which interact with nuclei and electrons in the atmosphere and produce secondary cosmic rays, consisting mainly of mesons, protons, neutrons, electrons, and photons of lower energy. *See* MESON; PROTON

All ionizing radiations produce biological changes (*see* RADIATION BIOLOGY), directly by ionization or excitation of the atoms in the molecules of biological entities, such as in chromosomes, or indirectly by the formation of active radicals or deleterious agents, through ionization and excitation, in the medium surrounding the biological entities Ionizing radiation, having high penetrating power, can reach the most vulnerable part of a cell ...

[illegible]

... depends in fact on the density of ionization (specific ionization or linear energy transfer, LET) along the path of the ionizing particle rather than on the nature of the particle It depends also on many other factors.

Use in medicine. The medical uses of nuclear radiations may be divided into three distinct classes:

1 The radiations, which are principally x-rays, are used to study the *anatomical* configuration of body organs, usually for the purpose of detecting abnormalities as an aid in diagnosis.

2. The radiations are used for therapeutic purposes to produce biological changes in such tissues as tumors

3 The radiations are used as a simple means of tracing a suitable radioactive substance through different steps in its course through the body, in the study of some particular physiological process. *See* RADIOLOGY. [G F.]

Use in biological research. The radiations emitted by radioactive isotopes of the various elements are used in biological research. The most useful ones in biological research are the isotopes of the elements which are important in metabolism and in the structural materials of cells. These include carbon, hydrogen, sulfur, and phosphorus. Unfortunately, nitrogen and oxygen do *not* have usable radioisotopes Isotopes of calcium, iodine, potassium, sodium, iron, and a few others have more limited usefulness in biological research In addition, the radioactive metals like cobalt 60, radium, and others can be used to produce radiations for external application to cells and tissues.

Most of the isotopes mentioned emit β-particles when they decay, and a few emit γ-rays. Therefore, they can be easily detected by various means If the radiations emitted are highly penetrating, like γ-rays and the high-energy β-particles from phosphorus-32, the presence of the isotope may be detected with Geiger counters or scintillation counters applied to the surface of the biological material Likewise, application of photographic emulsions to the surface or to cut surfaces of cells or tissues may serve to locate the isotope (*see* AUTORADIOGRAPHY) When the biological material may be broken down and destroyed, particular compounds or larger components of cells may be isolated, and the isotope determined by the various types of radiation detectors These procedures are sometimes used in the study of the movement of elements or their compounds in plants and animals They are frequently used for tracing the sequence of reactions in metabolism. The great advantage of radioisotopes for such studies is that small amounts or concentrations may be readily detected and measured, either in the presence of other isotopes of the same element or when mixed with a variety of other elements

In addition to the radiations emitted, some of the elements change to a different element when they decay Phosphorus-32 changes to sulfur; sulfur-35 changes to chlorine; and tritium (hydrogen-3) changes to helium when they decay by the emission of an electron, a β-particle Therefore, in addition to the radiation produced, the transmutation of the element affects the molecule and the cell of which it is a part. Attempts have been made to evaluate the effects of the two factors by giving the fungus *Neurospora* the same amount of radiation, and by having different propor... [illegible]

... of the radioisotope in two cultures, but in one, the radiosulfur or radiophosphorus was diluted greatly with the nonradioactive isotope. A difference in lethal effect could be demonstrated.

craft involves a consideration of physical size, which in many cases is more important than the economic factors associated with stationary plants. This is because the weight of the shield increases with the diameter of the reactor core. Because even the most compact cores require 50-100 tons of shielding, the problem of transporting this heavy shield is serious. Nuclear systems with as high a power density (power output per unit volume) as possible are, therefore, desired for mobile power reactors. Nuclear propulsion offers the advantages of lower-weight fuel, and less frequent fueling. See NUCLEAR AIRCRAFT PROPULSION; NUCLEAR ROCKET; REACTOR, SHIP PROPULSION.

Miscellaneous reactor applications. The heat and radiations from the fission reaction can be utilized in a number of ways other than for electricity generation and propulsion. Reactor-produced steam can be used for chemical processes or for heating buildings. Reactor heat and radiation are capable of promoting chemical reactions. Radioactive gamma emitters produced in the reactor can sterilize food and other products. Although such applications are all technically practicable, their extensive use will depend on costs of nuclear compared to conventional processes. For a discussion of reactor types, see REACTOR, NUCLEAR (CLASSIFICATION). [J.A.L.]

Bibliography: C. F. Bonilla (ed.), *Nuclear Engineering*, 1957; H. Etherington (ed.), *Nuclear Engineering Handbook*, 1958; S. Glasstone, *Principles of Nuclear Reactor Engineering*, 1955; D. B. Hoisington, *Nucleonics Fundamentals*, 1959; R. L. Murray, *Introduction to Nuclear Engineering*, 1954; R. Stephenson, *Introduction to Nuclear Engineering*, 2d ed., 1958.

Nuclear radiation

A term used to denote all the particles and radiations which emanate from the atomic nucleus as a result of radioactive decay and nuclear reactions. The term was originally used to denote only the ionizing radiations observed from naturally radioactive materials. These were α-rays (high-speed helium nuclei), β-rays (negative electrons or negatrons), and γ-rays (electromagnetic radiation of much shorter wavelength than visible light).

The distinction between nuclear radiations and others with similar physical properties lies in whether a nuclear process is involved in their production. Thus, although γ-rays and x-rays are both electromagnetic radiations and, for the same wavelength, are not distinguishable physically, one (γ-radiation) is emitted as a result of a rearrangement of protons and neutrons within the nucleus, while the other (x-rays) results from rearrangement of electrons outside the nucleus.

In addition to α-, β-, and γ-rays, other commonly encountered nuclear radiations are positively charged electrons (positrons), protons, and neutrons. Another radioactive decay product is the neutrino; it is not ordinarily considered as nuclear radiation, since its interaction with matter is very slight. Nuclear reactors are excellent sources of neutrinos. See NEUTRINO; NUCLEAR REACTION. [W.W.P.]

Nuclear radiation (biology)

Nuclear radiations are used in biology because of their common property of ionizing matter. This makes their detection relatively simple, or makes possible the production of biological effects in any living cell. Nuclear radiations originate in atomic nuclei, either spontaneously, as in radioactive substances, or through interactions with neutrons, photons, and so on. Gamma radiation originates in atomic nuclei and constitutes one kind of nuclear radiation, but it is otherwise indistinguishable in its effects from x-radiation produced by extranuclear reactions. Because x-rays have been readily available for many years, they have been used more extensively in biology and medicine than γ-rays. Therefore, x-rays must be included in any discussion of the biological and medical uses of nuclear radiations. See GAMMA RAYS; NEUTRON; PHOTON; X-RAY(S), PHYSICAL NATURE OF.

Ionizing radiation. Ionizing radiation is any electromagnetic or particulate radiation capable of producing ions, directly or indirectly, in its passage through matter.

Electromagnetic radiations. X-rays and γ-rays are electromagnetic radiations, traveling at the speed of light as packages of energy called photons. They ionize indirectly by first ejecting electrons at high speed from the atoms with which they interact; these secondary electrons then produce most of the ionization associated with the primary radiation.

Particulate radiation. Fast neutrons are particulate radiation consisting of nuclear particles of mass number 1 and zero charge, traveling at high speed. They ionize indirectly, largely by setting in motion charged particles from the atomic nuclei with which they collide. Slow or thermal neutrons ionize indirectly by interacting with nuclei, in a process known as neutron capture, to produce ionizing radiation.

Alpha rays are particulate radiation consisting of helium nuclei traveling at high speed. Since they are charged particles, they ionize directly. Alpha particles are emitted spontaneously by some radioactive nuclides or may result from neutron capture; for example, neutron capture by boron-10 produces lithium-7 and an α-particle. The energy of α-particles emitted by radioactive substances is of the order of a few million electron volts (Mev), but α-particles of very much higher energy may be produced in cyclotrons or other particle accelerators from helium-ion beams. With such machines other ionizing particles of very high energy, such as protons, deuterons, and so on, may also be produced. See ALPHA RAYS.

Beta rays are particulate radiation consisting of electrons or positrons emitted from a nucleus during β-decay and traveling at high speed. Since they are charged particles, that is, − or +

change in the kinetic energy of the deuteron.) In each case, the interaction of the incident particle with the target nucleus results in the formation of a residual nucleus and an emitted particle. In the (d,d') reaction, in which the residual nucleus is the same as the target nucleus, the process is referred to as scattering, either elastic or inelastic, depending upon whether the residual nucleus is left in its ground state or in one of its various excited states. The other three reactions lead to the production of a residual nucleus different from the target nucleus and are examples of nuclear disintegrations or transmutations. In these cases, also, the residual nucleus may be formed in its ground state or in one of its excited states. If the latter situation occurs, the residual nucleus will subsequently emit the excitation energy in the form of γ radiation or, occasionally, electrons. The residual nucleus may also be a radioactive species, as in the case of N^{13} formed in the $C^{12}(d,n)$ reaction. In this case, the residual nucleus will undergo further transformations in accordance with its characteristic radioactive decay scheme.

Q value An important quantity for nuclear reactions, the Q value, is defined as the difference between the kinetic energy of the products and the kinetic energy of the original particles. It is the total kinetic energy released in a nuclear reaction. Reactions with a positive Q value are called exoergic ...

Q ...

tio ...

cleus is formed in its ground state, the Q values are: $C^{12}(d,d')C^{12}$, $Q = 0$, $C^{12}(d,p)C^{13}$, $Q = 2.72$ Mev; $C^{12}(d,n)N^{13}$, $Q = -0.28$ Mev, $C^{12}(d,\alpha)B^{10}$, $Q = -1.39$ Mev. For reactions with a negative Q, a definite minimum energy, or threshold energy, is necessary for the reaction to take place. While there is no threshold energy for positive Q reactions, the yields of those reactions involving charged incident particles are quite low unless the bombarding energy is high enough to enable the incident ...

... values are numerically equal but have opposite signs. Thus, the Q for the $B^{10}(\alpha,d)C^{12}$ reaction is $+1.39$ Mev.

Conservation laws The probability that a particular reaction will take place when an individual target nucleus interacts with an incident particle is a function of the bombarding energy and the factors which determine it are not completely understood. However, it has been found experimentally that certain physical quantities are conserved in all nuclear reactions, and these conservation laws restrict the reactions which may take place. Those quantities which are conserved are described in the following paragraphs.

Charge. The total electric charge is always conserved. Except for high energy reactions involving meson production, the total number of protons is also conserved. In the $C^{12}(d,\alpha)B^{10}$ reaction, for example, there are seven protons involved in both the initial components and the final products of the reaction.

Mass number. The total number of nucleons is always the same both before and after the reaction. For each of the four reactions listed, 14 nucleons are involved. Since, except for reactions which result in meson production, the number of protons is conserved, the number of neutrons is also constant at each stage in nuclear reactions.

Energy. The total energy is conserved in all nuclear reactions, although neither the kinetic nor the rest energies are separately conserved. The conservation of total energy is expressed in the relation

$$m_1c^2 + T_1 + m_2c^2 + T_2 = m_3c^2 + T_3 + m_4c^2 + T_4$$

In this equation, the subscripts 1, 2, 3, and 4 refer to the incident particle, the target nucleus, the residual nucleus, and the emitted particle, respectively. The m's represent the rest masses of the neutral atoms, the T's are their kinetic energies, and c is the velocity of light. In the common experimental situation, T_2 is so small as to be negligible. In this equation, the kinetic energies and the rest masses are usually expressed in units of millions of electron volts, the conversion factor between the two being 1 atomic mass unit (amu) = 931.162 Mev.

Linear momentum. The total linear momentum is the same before and after any nuclear reaction. A consequence of this conservation law is that the threshold energy necessary to initiate an endoergic reaction is not numerically equal to the negative Q value, but is higher by the amount required to enable the final ...

mome ...

by th ...

the $C^{12}(d,n)N^{13}$ reaction, for example, is 0.33 Mev.

Angular momentum. The total angular momentum in nuclear reactions is the sum of the angular momentum associated with the relative motion of the reaction components and their intrinsic angular momentum, or spin. This total is always conserved.

Parity. Experimental evidence shows that, in most nuclear reactions, the total parity is the same before and after the interaction. Since the parity associated with the wave function describing the motion of a particle is determined by the ...

... of its allowed states has either even or odd parity, this conservation law, together with that for angular momentum, acts to restrict those excited states of the residual nucleus which can be formed by an incident particle of given angular momentum. See PARITY (QUANTUM MECHANICS)

Statistics. Since the total number of nucleons is conserved during a nuclear reaction, the statistics which govern the system are the same before, during, and after the interaction. Fermi-Dirac statis-

In other experiments, the decay of phosphorus-32 has been used to give information on the nature and importance of the phosphorus-containing molecules to the survival or reproduction of a cell or virus particle. When bacterial viruses are allowed to grow in the presence of phosphate containing phosphorus-32, they incorporate the radioisotope into their genetic material, deoxyribonucleic acid (DNA). If these virus particles are then stored under conditions where no growth occurs, the decay of the radioisotope is very effective in inactivating the viruses. About 1 in 10 atoms which decay inactivates 1 virus particle. From such experiments, biologists are able to learn something about the importance of the molecule as well as something of its size and possible organization. Similar experiments have been carried out with the cells of ameba, but the lethal effect was observed, not on the stored cells, but on their offspring. From this experiment, deductions concerning the organization of genetic material and its mutability were drawn. *See* DEOXYRIBONUCLEIC ACID.

Nuclear radiations have also proven useful in many studies of the nature and the mechanisms of the effects of radiations on cells and cell constituents. The radiations with low penetrating power, for example, α-particles and low-energy β-particles, can be most effective when an element which will emit the particles is placed inside the cells to be studied. Radon, which is a gas, can be used in this way for the production of α-particles. Likewise, a variety of the heavy metals like thorium, uranium, and polonium emit α-particles. Various β-emitters, such as phosphorus-32, carbon-14, sulfur-35, and tritium can also be used for producing radiations inside the cell. One of the most interesting of this group is tritium which emits very soft β-particles. Their maximum range is 6 μ in water and about the same in tissues, but the average range is much less, about 1 μ or 1/25 the diameter of a medium-sized cell. Tritium can be put into the cell as water or in many other compounds. Many of these are unselective and are equally distributed to all parts of the cell. However, there is one substance, thymidine, which selectively labels the DNA which is restricted to the cell nucleus. If the cell is a relatively large one compared to its nuclear size, nearly all of the radiation will be absorbed by the nucleus, while the other part of the cell is irradiated hardly at all. Experiments have shown that tritium given to cells in the form of tritium-thymidine is about 1000 times as effective as tritium-water in producing radiation effects. *See* HISTORADIOGRAPHY; SCINTILLATION COUNTER. [J.H.T.]

Nuclear reaction

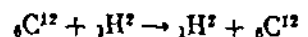
A reaction which is produced as a result of interactions between atomic nuclei when the interacting particles approach each other to within distances of the order of nuclear dimensions (10^{-12} cm). In the usual experimental situation, one of the interacting particles, the target nucleus, is essentially at rest, and the reaction is initiated by bombarding it with nuclear projectiles of some type.

Means of producing reactions. Because of the intense electrostatic field produced by the nuclear charge, positively charged bombarding particles must have a large kinetic energy in order to overcome the electrostatic (Coulomb) repulsion and reach the target nucleus. For this reason, the ions most commonly used are those of the isotopes of hydrogen (protons, deuterons, and tritons) and helium (He^3 and He^4). While for the lightest target nuclei, protons with kinetic energies of a few hundred thousand electron volts are sufficient to cause certain reactions, energies of 10–15 million electron volts (Mev) are required for heavy target nuclei. Beams of such energetic charged particles are provided by particle accelerators of various types (Van de Graaff generators, cyclotrons, linear accelerators, and so forth). For information on particle-beam production, *see* PARTICLE ACCELERATOR.

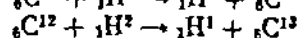
Since neutrons are uncharged, they are not repelled by the electrostatic field of the target nucleus, and neutron energies of only a fraction of an electron volt are sufficient to initiate nuclear reactions. Neutrons for reaction studies may be obtained from nuclear reactors or from various nuclear reactions. The interaction of electromagnetic radiation with nuclei may also lead to nuclear reactions. So-called photodisintegration may take place if the radiation has sufficient energy to cause the target nucleus to break up into two or more fragments. In a similar manner, high-energy electrons may also cause nuclear disintegrations. Electromagnetic radiation and electrons, however, interact strongly with the atomic electrons surrounding the target nucleus and are relatively less effective in causing nuclear reactions than are nuclear particles such as protons and neutrons.

The most common and most extensively studied reactions are those which result in two products, one of which, the residual nucleus, is of nearly the same mass number and charge as the target nucleus, while the other product, the emitted particle, is either a single nucleon (a proton or a neutron), or a small assembly of nucleons, such as an α-particle. If the bombarding energy is sufficiently high, a spallation reaction occurs in which three or more products may result. *See* SPALLATION REACTION.

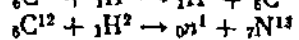
Typical reactions. Four common types of nuclear reactions are observed when a layer of carbon-12 is bombarded with deuterons. Deuterons, protons, neutrons, and α-particles are emitted, the following reactions being responsible:



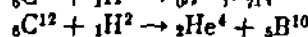
$$_6C^{12} + {}_1H^2 \rightarrow {}_1H^2 + {}_6C^{12}$$



$$_6C^{12} + {}_1H^2 \rightarrow {}_1H^1 + {}_6C^{13}$$



$$_6C^{12} + {}_1H^2 \rightarrow {}_0n^1 + {}_7N^{13}$$



$$_6C^{12} + {}_1H^2 \rightarrow {}_2He^4 + {}_5B^{10}$$

These reactions are conventionally written as $C^{12}(d,d')C^{12}$, $C^{12}(d,p)C^{13}$, $C^{12}(d,n)N^{13}$, and $C^{12}(d,\alpha)B^{10}$, respectively. (T 'me indicat

occur over a wide range of proton energies If the compound nucleus decays to its ground state by the emission of a γ-ray, the process is known as a (p,γ) reaction With higher proton energies, a (p,n) reaction is possible. This always has a negative Q value and leads to a radioactive residual nucleus For many target nuclei, the (p,α) reaction has a high positive Q, but the yields are low, except at high proton energies, because of the difficulty of the doubly charged α-particle in penetrating the nuclear barrier

Deuteron-induced reactions The (d,p), (d,n), and (d,α) reactions usually have positive Q values Except for light nuclei, where the nuclear potential barrier is low, the (d,α) reactions have low probabilities The (d,n) reactions of deuterium, tritium, and beryllium are important as sources of neutrons Both the (d,p) and (d,n) reactions often lead to radioactive residual nuclei that are useful in various fields of investigation

Neutron-induced reactions Neutron capture leading to an (n,γ) reaction is important for all stable nuclei, and occurs even with very low energy neutrons With a given target nucleus, it yields the same final product as the (d,p) reaction The capture γ-rays usually have maximum energies of about 8 Mev This reaction is the source of many of the radioactive isotopes produced by nuclear reactors For high-energy neutrons, the (n,p) and (n,α) reactions are also observed In very heavy nuclei, neutron capture may lead to disintegration of the compound nucleus into two massive fragments, with the release of large amounts of kinetic energy ...

... tions of various light nuclei using the α-particles from naturally occurring radioactive substances were the first examples of artificially produced nuclear disintegrations High α particle energies are required for other than light nuclei because of the Coulomb barrier of the nucleus At sufficiently high energies (about 30 Mev), (α,p) and (α,n) reactions are observed, even for heavy nuclei. See RADIOACTIVITY, TERRESTRIAL NUCLEAR REACTIONS, *see also* FUSION, NUCLEAR, THERMONUCLEAR REACTION.

[W.W.BU.]

Bibliography R D Evans, *The Atomic Nucleus*, 1955, S Fluegge (ed), *Handbuch der Physik*, vol 40, 1957

Nuclear rocket

A type of thermal engine utilizing nuclear fission or fusion reactions to heat a working fluid for propulsive purposes

The heat-transfer nuclear rocket employs fixed fuel elements containing uranium-235 or plutonium 239 Graphite, tungsten, and some of the carbides make good rocket reactor ... as high as 6000°R. The hot gases are expanded through a nozzle to produce thrust. Specific impulses up to 900 seconds are theoretically possible with high-pressure, high-thrust nuclear rockets Thrust levels for some applications are on the order of 1,000,000 lb. Payload-to-launching weight ratios up to 0.1 seem possible. Low-pressure (0 01 atm) hydrogen, heat-transfer rockets may attain specific impulses of approximately 1500 sec in outer space.

The barrier problems of the heat-transfer nuclear rockets include the maintenance of fuel element integrity at high temperatures for the engine lifetime, release of radioactivity, complexity compared to chemical rockets, and cost of fissionable fuel and propellants. For booster applications, engine life may be only several minutes (*see* ROCKET STAGING). The U.S. Atomic Energy Commission is developing these rocket concepts as part of the Rover project (*see* NUCLEAR AIRCRAFT PROPULSION; SPECIFIC IMPULSE; THRUST).

Consumable, controlled explosion, or plasma core nuclear rockets isolate the reactor core from the temperature-limited solid structure of the rocket by electromagnetic fields or fluid vortices. In this approach, the fission reaction occurs directly in the working fluid, leading to propellant temperatures up to 100,000°R. Specific impulses may approach 2500 sec in theory Thermal radiation to the solid structure, the loss of expensive unfissioned fuel, and the expulsion of radioactive fission products to the surroundings severely limit the potential of this type of rocket. Nuclear bomb propulsion utilizes the blast effects of small, repeated atomic explosions to propel space vehicles. High instantaneous accelerations, loss of fissionable fuel, and radioactive hazards make this a questionable type of propulsion system.

Thermonuclear rockets use the fusion reactions to heat a working fluid. Deuterium, tritium, and lithium are possible fuels. In conceptual designs the high-temperature (10^{8}°K) reaction is confined by electromagnetic fields. Heat is transferred to the surrounding propellant by thermal radiation and high-kinetic-energy neutrons The problem of containing the high-temperature plasma has not yet been solved *See* PLASMA PHYSICS, PROPULSION.

[W.R C.]

Nuclear structure

The atomic nucleus is at the center of the atom and contains all except 0 02% of the mass of the atom. Its density is 4×10^{8} tons/in.^{3}, and its diameter is about 10^{-12} cm.

The nucleus is positively charged and balances the negative charge of the electrons of the atom, thus making the atom as a whole neutral The positive charge resides on the Z protons (Z = atomic number) in the nucleus. The proton itself is the nucleus of the lightest isotope of hydrogen (ordinary hydrogen, H^1). But the mass of most nuclei is greater than the mass of the Z protons they contain by a factor of 2 or more The difference is made up

tics are obeyed if the total number of nucleons is odd, and Bose-Einstein if the total number is even. *See* QUANTUM STATISTICS.

Reaction mechanisms. A number of mechanisms have been proposed to account for the observed features of nuclear reactions. While none have been completely successful, they provide means for correlating and at least partially understanding many of the experimental facts. The most generally used models for nuclear reactions are described here.

Compound nucleus formation. According to this point of view, originally proposed by N. Bohr, a nuclear reaction is visualized as proceeding in two distinct steps. The incident particle and the target nucleus are assumed to combine to form a compound nucleus, which exists for a time (of the order of 10^{-16} sec) which is much longer than the approximately 10^{-22} sec that would be required for the incident particle to pass through the target nucleus. The compound nucleus is always in a highly excited, unstable state and can subsequently decay into a number of different products, or through a number of so-called exit channels. In the four examples cited earlier, $_7N^{14}$ is the compound nucleus formed by the amalgamation of a deuteron and $_6C^{12}$, and four possible decay modes or exit channels are indicated. Two essential features of this hypothesis are that, during its relatively long lifetime, the compound nucleus "forgets" the particular way in which it was formed, and that the energy brought in by the incident particle is shared by all the nuclear constituents. The probability that a particular reaction will occur is, then, the product of the probability of forming the compound nucleus and the probability that it will decay through a particular exit channel. Experiments indicate that, for a given energy of excitation in the compound nucleus, this latter factor is independent of the manner in which the compound nucleus is formed. In the case of N^{14}, it can be formed by $C^{13} + p$ or $B^{10} + \alpha$, as well as by $C^{12} + d$. While certain features of various types of interactions cannot be completely explained on the compound nucleus hypothesis, it appears that this mechanism plays some role in nearly all nuclear reactions.

Direct interactions Some reactions have probabilities or other properties which conflict with the predictions of the compound nucleus hypothesis, and many are better explained on the assumption that the incident particle does not combine with the target nucleus as a whole, but rather that it, or some component, interacts only with the surface or with some individual constituent. The entire process is completed in the time required for the ... the diameter of the ... itted par- ... distribu- tions favoring the forward direction tend to be of this type.

The observed properties of most (d,p) and (d,n) reactions are consistent with the assumption that, in the passage of a deuteron near a target nucleus, the neutron and proton are separated because of the nuclear electric field and that only one or the other is captured, the other being "stripped off" to emerge as the emitted particle of the reaction. This mechanism was first proposed by J. R. Oppenheimer and M. Phillips to account for the relatively high probability of (d,p) reactions. The inverse mechanism, in which an incident particle captures or "picks up," a particle from the target nucleus apparently takes place in (p,d) and (d,H^3) reactions. Stripping and pickup reactions may be considered as examples of direct interactions.

Coulomb excitation. It is observed that, in the bombardment of nuclei with protons, deuterons, and α-particles, inelastic scattering, resulting in excitation of the target nucleus, occurs at bombarding energies so low that the probability of either direct interaction or compound nucleus formation is negligible. This process is well explained by the assumption that the nuclei interact with the rapidly changing electric field caused by the passage of the charged bombarding particle.

Elastic scattering. This process leaves the quantum state of the scatterer unchanged For charged bombarding particles with low energies, the elastic nuclear scattering is accurately described in terms of the inverse-square force law between electric charges. In this case, the process is known as Rutherford scattering. For higher bombarding energies, where the particle can come within the range of the various nuclear forces (approximately 10^{-13} cm), the scattering deviates from predictions based on the inverse-square law. In the case of neutrons, the elastic scattering is entirely due to the nuclear forces. For a discussion of nuclear forces, *see* NUCLEAR STRUCTURE; *see also* SCATTERING EXPERIMENTS, NUCLEAR.

Nuclear cross sections. The cross section for a nuclear reaction is a measure of its probability Consider a reaction initiated by a beam of particles bombarding a region which contains N atoms per unit area (uniformly distributed) and where I particles per second striking the area result in R reactions of a particular type per second. This result can be expressed in terms of the fraction of the bombarded region which is effective in producing reaction products, R/I. If this is divided by the number of nuclei per unit area, the effective area or cross section per target nucleus is obtained The cross section $\sigma = R/IN$ This is referred to as the total cross section, since it involves all the disintegration products of the reaction The dimensions are those of an area, and total cross sections are expressed in either square centimeters or in barns (1 barn = 10^{-24} cm^2).

Types of reactions. Aside from elastic and inelastic scattering, the most common interactions initiated by the usual bombarding particles are discussed in the following paragraphs

Proton induced reactions. Capture reactions, in which the proton combines with the target nucleus to form a compound nucleus in an excit ... e,

(2) the extra electrostatic energy of the proton-rich nucleus. Contribution (1) is accurately known. Contribution (2) would be

$$\frac{6}{5}\frac{Ze^2}{R}$$

where e is the magnitude of the charge on the electron or proton, if the nucleus behaved classically and were a sphere of radius R. In practice this last assumption is too crude, but the appropriate quantum-mechanical corrections can be made. This mass difference manifests itself in the instability of the heavier (proton-rich) nucleus which transforms itself radioactively into the lighter (neutron rich) nucleus usually by the emission of a positron, although K-capture (the taking into the nucleus of an orbital electron of the atom) is sometimes found. The energy of this transition can then be related to R, the nuclear radius.

Other methods. Many other methods for finding nuclear radii exist, for example, the scattering by nuclei of strongly interacting particles such as neutrons, protons, or π-mesons; α-radioactivity; and the so-called fine structure of atomic and x-ray spectra. *See* FINE STRUCTURE (SPECTRAL LINES), SCATTERING EXPERIMENTS, NUCLEAR.

Nuclear densities. The methods just described give concordant results. From them (especially from the scattering of fast electrons) it has been determined that the nuclear density distribution can be quite accurately described by the formula

$$\rho(r) = \rho(0)\{1 + e^{(r-R)/a}\}^{-1}$$

where $\rho(r)$ is the density at a distance r from the center and $\rho(0)$ is the density at the center. This is the bell-shaped curve which is drawn in Fig. 1 for a nucleus of gold. The radius R is the point at which the nuclear density has fallen to one half of its central value. Experiments give

$$R = 1.07\ A^{1/3} \times 10^{-13}\ \text{cm}$$

The a in the equation for the nuclear density does not change appreciably with A and is such that the surface thickness t, the distance over which the nuclear density falls from 90 to 10% of its central value, is

$$t = 2.3 \times 10^{-13}\ \text{cm}$$

It is important to note (1) that the nucleus is not uniform in density but has an appreciable region of gradual change of density, and (2) that the radius R is proportional to $A^{1/3}$, which implies that the volume is proportional to A and thus that the mean density is independent of the size of the nucleus. Phenomenon (2) is referred to as saturation.

Nuclear masses. The mass of the nucleus gives a great deal of important information about nuclear structure. Nuclear masses can be measured using mass spectrometers (*see* MASS SPECTROSCOPE) or by study of nuclear disintegrations. The

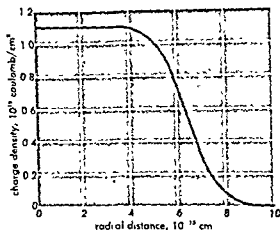


Fig. 1. Radial distribution of charge in a nucleus of gold.

unit of nuclear mass most commonly used is equal to 1/16 of the mass of the atom of oxygen of $A = 16$.

The mass $M(A,Z)$ of an atom can now be written. The first term is just the mass of the Z hydrogen atoms plus the mass of the $N = A - Z$ neu-

nucleus with an energy ...

pendent of their number, as is evidenced by the more or less constant nuclear density. This contribution is then $-m_1A$, where m_1 is a constant. However, the nucleons on the surface of the nucleus feel only inward attractions and thus are not so strongly bound. This gives rise to a surface-tension term in the mass formula proportional to the surface area, namely, $+m_2A^{2/3}$ (because $R \sim A^{1/3}$).

One should a priori expect nuclei to contain equal numbers of neutrons and protons, because successive particles of the same kind must go into successively higher quantum states (the Pauli exclusion principle forbids two identical particles to share the same state; *see* EXCLUSION PRINCIPLE) so the lowest total energy is found when the number of neutrons and protons is equal. If there were, say, more neutrons than protons, then the neutrons would be in higher energy states than the protons and thus it would be energetically profitable for some neutrons to convert themselves to protons by the radioactive emission of electrons. The inequality of neutron and proton numbers gives rise to the mass term $+m_3(N - Z)^2/A$. However, the energy of the protons that is due to their purely electrostatic mutual interaction gives a greater energy

diminish the proton number and increase the neutron number, the higher quantum state of the neutron not requiring as much extra energy as is

by neutrons. The neutron is a neutral particle which is 0.14% more massive than the proton. The total number of neutrons and protons in a nucleus is the mass number A. *See* ATOMIC STRUCTURE AND SPECTRA; ELECTRON; NEUTRON; PROTON.

It was thought at one time that the nucleus got its mass from A protons but that it also contained $A - Z$ electrons to diminish its positive charge to Z. This theory is now known to be incorrect for three reasons.

1. If the nucleus contained electrons, they would have a momentum p which would be related to the nuclear radius R through the Heisenberg uncertainty relation $pR \sim \hbar$, where $\hbar$ is Planck's constant h divided by 2π (*see* UNCERTAINTY PRINCIPLE). This momentum corresponds to an electron energy of about 20 Mev. However, there is no force of any strength other than the Coulomb electrostatic attraction operating between protons and electrons, and this is not nearly strong enough to hold within the small confines of the nucleus electrons of this energy.

2. Protons and electrons have an intrinsic angular momentum (spin) of $\frac{1}{2}\hbar$. Consider the nucleus N^{14} (nitrogen-14). If it contained 14 protons and 7 electrons, the total number of particles each of spin $\frac{1}{2}\hbar$ would be odd and so the total angular momentum of the nucleus would have to be $\frac{1}{2}\hbar$, $\frac{3}{2}\hbar$, ... Experimentally, the total angular momentum is found to be $\hbar$. This is consistent with the vector addition of the intrinsic spins of 7 protons and 7 neutrons (the intrinsic spin of a neutron is also $\frac{1}{2}\hbar$). There are many such examples. If A is even, the total nuclear angular momentum is always an integral multiple of $\hbar$, if A is odd, it is always an odd number of half-integral units.

3 The magnetic moment of the electron (Bohr magneton) is roughly 1000 times greater than that of the neutron and proton. The magnetic moments of nuclei are about the same size as those of the neutron and proton, thus nuclei cannot contain electrons. *See* MAGNETON; SPIN (QUANTUM MECHANICS).

Although electrons cannot be permanent elements of nuclear structure, they are found to be emitted at high speed by radioactive nuclei—the phenomenon of β-decay or β-radioactivity. Both negative-electron and positive-electron (positron) emission are known. This phenomenon results from the fact that the neutron and proton can decay into each other radioactively with the emission of positive or negative electrons. The method of decay is determined by the masses of the atoms concerned. A free neutron with a half-life of 12 min decays into a proton and a negative electron. Together with the electron or positron, an antineutrino or neutrino is emitted—this particle has no rest mass and has intrinsic spin $\frac{1}{2}\hbar$ See NEUTRINO; RADIOACTIVITY.

Because the neutron and proton have almost the same mass, the same spin, the same interactions, and can interconvert in the phenomenon of β-decay, they are regarded not so much as being separate particles but merely as different states of a single particle, the nucleon, distinguished only by the charge which acts as a sort of label

Nuclear sizes. The size of the nucleus is an important parameter. It can be determined in many ways, of which the most accurate are the scattering of fast electrons, measurement of energies of x-rays emitted by μ-mesons, and study of mirror nuclei.

Scattering of fast electrons. When a high velocity electron approaches a nucleus, it is deflected from its path because of the electrostatic attraction of the positively charged nucleus This deflection can be accurately calculated. If the nucleus is of finite size and the encounter is close, the electron will penetrate the nucleus; it will then feel a smaller attraction than for a point nucleus and so its deflection will be less and will depend on the nuclear size. If the encounter is not a close one the electron will not penetrate the nucleus, will always feel the same attraction as for a point nucleus, and the deflection will not depend on the nuclear size Thus from a comparison of the probabilities of large and small deflections, physicists can determine the size of the nucleus. The deflections so measured with electrons of up to 500 Mev energy from a linear electron accelerator are entirely explained on the basis of the Coulomb force alone. There is no evidence for any strong nonelectric force between electrons and nucleons.

Mesonic x-rays When a negative μ-meson (of mass $m_\mu = 207m_e$, where m_e is the mass of the electron) is brought to rest in matter, it is attracted by the positive charge of a nearby nucleus and circles around it in orbits just as an electron does It makes transitions between these orbits, emitting x-rays as it goes. Because its mass is so much greater than that of an electron, the orbits are correspondingly smaller (the orbit sizes vary inversely as the mass) and the x-ray energies are correspondingly higher (the energies are proportional to the mass). The lowest orbits are actually smaller than the nucleus for the heavier elements The motion of the μ-meson is therefore greatly modified from what it would have been for a point nucleus, the energy levels and thus the x-ray energies are altered in a predictable manner depending upon the size of the nucleus

Mirror nuclei. The interactions of nucleons with each other do not depend on whether they are neutrons or protons (apart from the Coulomb force in the case of interacting protons); thus it is logical to assume that a nucleus containing x neutrons and y protons will differ from one containing y neutrons and x protons only through the Coulomb force Such related nuclei are called mirror nuclei and are discussed in detail later Typical examples are C^{13} and N^{13} ($x = 6$, $y = 7$) In particular it is expected that the mass difference between the ground states of such mirror nuclei should be due to (1) the neutron-proton mass difference and

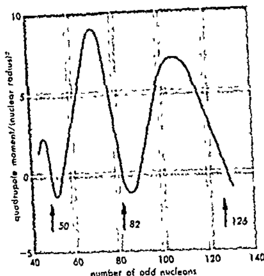


Fig 5 Plot of nuclear quadrupole moments for nuclei of odd A as a function of the number of odd nucleons. Note the low values at the magic numbers 50, 82, 126

even stronger Coulomb barrier than does α-particle *emission* and becomes important only for the relatively new artificially produced elements in the region of $Z = 100$ *See* BINDING ENERGY, NUCLEAR, FISSION, NUCLEAR, POTENTIAL BARRIER.

Magic numbers; nuclear shapes. The semiempirical mass formula predicts a smooth dependence of nuclear mass on Z and A In practice, striking discontinuities are observed at *certain values of* Z and N, namely 8, 20, 28, 50 82, and 126 One of these is illustrated in Fig 3, which shows the break in the binding energy of the last neutron in various elements as N passes through 126 These are the so called

many

shows

neutron ... Mev in many elements Striking minima are seen at the magic numbers Figure 5 shows the nuclear quadrupole moments which vanish at the magic numbers

The quadrupole moment is a measure of the departure of the nuclear charge distribution from sphericity It is positive for a football like *deformation*, and *negative* for a coinlike deformation. Figure 5 shows that nuclei are spherical at the magic numbers and ellipsoidal in between The large quadrupole moments seen around 70 on the abscissa scale in Fig 5 belong to the rare earths and correspond to a football like deformation with the length about 30% greater than the width. *See* NUCLEAR MOMENTS

Shell model. The shell model of the nucleus provides an explanation of the *magic* numbers Nuclei seem to behave as though they were attractive potential wells within which the nucleons move in orbits, just as do electrons around an atom The potential wells look very much like the nuclear matter distribution of Fig 1 and are about 50 Mev deep. Each nucleon moves with its *intrinsic spin* of $\frac{1}{2}\hbar$ coupled onto its orbital angular momentum $l\hbar$ to make a total angular momentum $j\hbar = (l \pm \frac{1}{2})\hbar$. These individual js finally couple together to make the total nuclear angular momentum $J\hbar$ (see QUANTUM NUMBERS). This is the jj-coupling model. If one assumes that a nucleon of $j = l + \frac{1}{2}$ is more tightly bound in the potential well than one of $j = l - \frac{1}{2}$, one can account for the magic numbers as natural breaks in the neutron or proton level scheme. This is shown in Fig. 6. One level or shell will hold $2j + 1$ *nucleons corresponding to* the $2j + 1$ allowable orientations of j. Nuclei which contain magic numbers of both neutrons and *protons, such as* Pb^{208} ($Z = 82$; $N = 126$), are called doubly magic and possess unusual stability.

Nuclear spins and parity. This jj scheme can also predict the spins (J values) and parities of nuclear ground (lowest energy) states. Parity is a strictly quantum-mechanical concept which arises because a particle such as a nucleon has also a wavelike aspect and can, for example, be diffracted This means that a nucleon must be described by an amplitude whose square gives the local particle probability or density. Although the particle density must clearly be the *same at a* given point in a nucleus and at an equal distance the other side of the center, the amplitude itself might or might not differ in sign at *those two points* If the *sign* is the same, the *parity* of the state is even (+), *if it is different*, the parity is *odd* (−). The parity of an individual nucleon moving in an orbit with angular momentum $l\hbar$ is $(-)^l$. *See* PARITY (QUANTUM MECHANICS).

If it *is* assumed that an even number of nucleons moving *in* the same shell will tend to pair off *their spins* to give zero resultant, the spin and parity of a nucleus of odd A can be predicted. The *spin* will just be the j *value* of the last nucleon as one fills the scheme of Fig 6 from the bottom, allowing $2j + 1$ *nucleons* per shell, the parity will be $(-)^l$ where l is the orbital angular *momentum* quantum *number* of the last (odd) nucleon. Separate schemes are filled *for* neutrons and protons. This procedure has had great success in accounting for the spins and

make predictions about odd odd nuclei, because it is not known how the j of the odd neutron should couple to that of the odd proton. The behavior is in fact *capricious*, and some remarkable spins *are* found, such as $J = 6$ for Lu^{176}. Even as light a nucleus as B^{10} has $J = 3$.

For light nuclei the shell model still holds but the jj-*coupling* scheme *is not* adequate. An alternative scheme is for all the orbital ls of the individual nucleons to couple together to form a grand L and all their intrinsic spins to *form S*, and *then for J* finally to be formed by coupling L and S. This is called the *LS*-coupling model, and this model may be valid for the very lightest nuclei. More often the situation is somewhere between

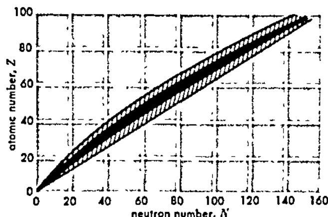


Fig. 2. The stability valley. Nuclei in the solid area are generally stable or are α-particle emitters of long lifetime; those in the shaded area are generally β-radioactive; those outside are unstable against neutron or proton emission, or against α-particle emission of short lifetime.

saved by diminishing the Coulomb energy of the protons. Light nuclei (up to about $A = 40$) tend to have roughly equal numbers of neutrons and protons, but *beyond this the relative neutron excess* grows until for the heaviest nuclei there are about 1.6 times as many neutrons as protons. The Coulomb energy contribution to the mass formula is $+m_4Z^2/A^{1/3}$.

Finally there is a term $+m_5$ which arises from the fact that neutrons and protons each have an intrinsic spin of $\frac{1}{2}\hbar$ and so the quantum energy states lie in pairs with the two like particles having their spins in opposite directions. An even number of like particles is thus favored because the next (odd) particle must go by itself into a higher quantum state. Thus the most stable nuclei are those containing even numbers of both neutrons and protons (even-even nuclei); next come those with an even number of neutrons and an odd number of protons or vice versa (even-odd or odd-even nuclei); least stable are the (odd-odd) nuclei with odd numbers of both neutrons and protons. Therefore the mass of an atom is

$$M(A,Z) = 1.00813\,Z + 1.00898\,N - m_1A + m_2A^{2/3} + m_3\frac{(N-Z)^2}{A} + m_4\frac{Z^2}{A^{1/3}} + m_5$$

This is called the semiempirical mass formula and the best values for the various constants give

$$M(A,Z) = 0.99391\,A - 0.00085\,Z + 0.014\,A^{2/3} + 0.021\,\frac{(N-Z)^2}{A} + 0.000627\,\frac{Z^2}{A^{1/3}} \pm 0.036\,A^{-3/4}$$

In the last term + is used for odd-odd nuclei, − for even-even nuclei. The term is struck out when A is odd.

Figure 2 shows as the cross-hatched area the so-called stability valley within which the stable nuclei are found. The shaded area around it indicates the region within which nuclei are *sufficiently stable to* be ordinarily radioactive and which approach the stability valley by β-decay. Beyond this region the instability is so great that nucleons are unbound and the nucleus rapidly enters the shaded area by shedding neutrons or protons.

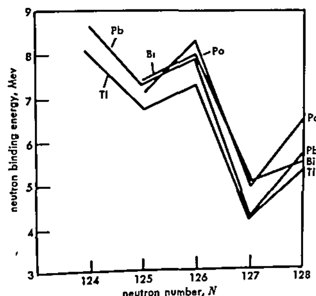


Fig 3. Binding energy of the last neutron in various elements as a function of neutron number N. Note the break after the magic number $N = 126$.

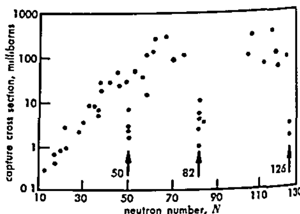


Fig. 4 Cross sections for radiative capture of neutrons of about 1 Mev in many nuclei as a function of neutron number N. Note the low values at the magic numbers $N = 50$, 82, and 126. One millibarn $= 10^{-27}$ cm^2

The most stable nuclei are around iron ($Z = 26$). Below that region the surface-tension term is relatively important and diminishes the binding, while above it the Coulomb term grows in importance and has the same effect. Emission of α-particles becomes energetically possible at about the middle of the periodic table. This emission is strongly discouraged by the electrostatic Coulomb potential barrier through which the α-particles have to penetrate, and with rare exceptions it becomes noticeable only for elements heavier than lead. Similarly, spontaneous fission of the nucleus into fragments of comparable mass is energetically possible over much of the periodic table, but fission fa n

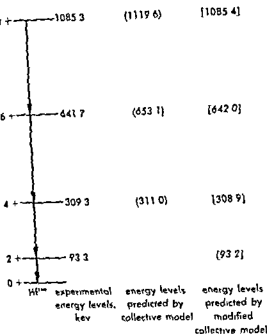


Fig 8 Experimental level scheme of Hf^{180} Symbols at left are the spins and parities, which agree with the theoretical values, as do the chains of γ-rays by which the levels deexcite, as shown

Nuclear forces. The nucleons which form the nucleus are held together by forces that are much *stronger than* the electrostatic forces that tend to disrupt the nucleus The nuclear force between two nucleons also differs from the Coulomb force in that it is short ranged, it drops off very rapidly after a certain distance, about 10^{-13} cm It is attractive at long distances but strongly repulsive at very short distances. The rough form of the *nucleon-nucleon* potential is shown in Fig 9 The radius of the repulsive core may be 0.4×10^{-13} cm The attractive part of the potential has the approximate form $e^{-r/a}/r$ The repulsive core is one of the two features that produce the saturation of the nuclear density that was referred to earlier in this article The other is the Pauli exclusion principle, which has the effect of keeping the nucleons apart

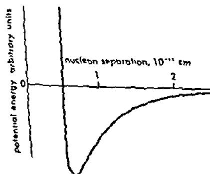


Fig 9 The potential operating between two nucleons due to the nuclear forces.

The nuclear force depends on the relative orientation of the two nucleon spins For a neutron and a proton, it is greater when the spins are parallel than when they are antiparallel. The force with the spins parallel and no orbital angular momentum is strong enough to bind the particles together to form the deuteron (binding energy 2.227 Mev); that with the spins antiparallel is not strong enough (*see* DEUTERON). The nuclear force between two neutrons is nearly the same as that between two protons. This is shown by comparing the level schemes of mirror nuclei such as have already been discussed. The only nuclear difference between such nuclei is that in going from the proton-rich to the neutron-rich nucleus, one exchanges certain proton-proton bonds for equivalent neutron-neutron bonds All the other bonds cancel out, as can be seen by counting them. Thus if these two interactions are the same, the level schemes of the two nuclei should be identical This is accurately borne out in practice (see Fig. 10).

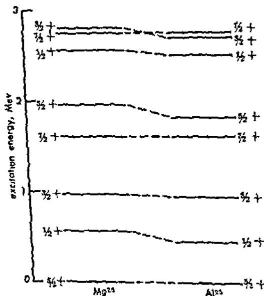


Fig. 10 Comparison of the experimental energy levels of the mirror nuclei Mg^{25} and Al^{25}. The ground states have been put at the same level in order to compare the ordering and excitation of the excited states. In practice, Al^{25} decays by positron emission into Mg^{25}. The comparison is carried only to the region where the experimental situation becomes uncertain The numerical values on the levels represent the spins

This equivalence is *called* the charge symmetry of nuclear forces A wider equivalence is the charge independence of nuclear forces, which asserts that the neutron-proton force is the same as the neutron-neutron or proton-proton force if the various pairs of particles are in similar spectroscopic states. This is also well tested by experiment. An immediate consequence is that the di-neutron cannot exist Two neutrons with their spins parallel and without orbital angular momentum are forbidden by the Pauli exclusion principle to interact,

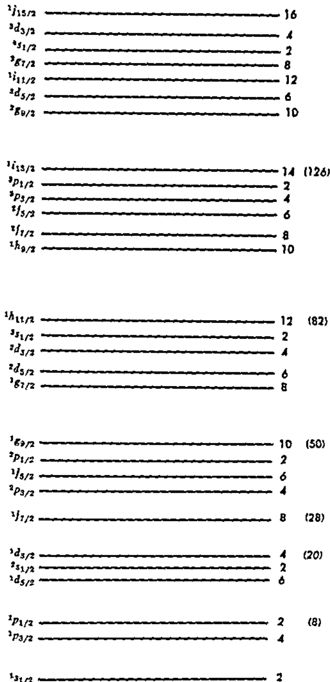


Fig. 6 Energy levels of nucleons according to shell model Symbols to left of levels indicate their spectroscopic character. Letter indicates orbital angular momentum ($s,p,d,f,g,$ . . . , for $l = 0,1,2,3,4,$ . . .); superscript indicates order of appearance of the various levels of the same orbital angular momentum, subscript gives j value Numbers to right of each level show how many neutrons or protons the level can hold $(2j + 1)$ Numbers further to the right in parentheses give total number of nucleons held up to and including that level Note the breaks at the magic numbers 8, 20, 28, 50, 82, and 126.

these schemes and is called intermediate coupling. Figure 7 shows a typical success that is achieved by the shell model in intermediate coupling for the excited states of a nucleus. In such a detailed calculation, allowance is made for the forces that operate between the several nucleons within one shell and give different energies for the different ways of coupling these nucleons together. Excited states can also be generated by raising a nucleon from one shell to a higher one.

Collective model. The collective model of the nucleus is suggested by some striking regularities in the level schemes of the heavy nuclei in the region of the rare earths, that were noted from Fig 5 as being heavily deformed. An example is shown in Fig. 8. These states are produced by the rotation of the deformed nucleus about a minor axis. Their expected excitation energies are given by

$$E = \frac{\hbar^2}{2I} J(J+1)$$

Where I is the effective moment of inertia of the nucleus; it is about one-half that of a rigid body of the same shape. For an even-even nucleus such as Hf^{180}, only the values $J = 0, 2, 4, \ldots,$ are theoretically allowed and experimentally found. As the nucleus spins faster it will tend to elongate itself even more because of the centrifugal forces, and so the moment of inertia tends to increase slightly with J This effect is noted in Fig. 8 by the terms in brackets, obtained by adding a theoretical term $\xi J^2(J+1)^2$ to the simple rotational spectrum. The coefficient ξ is a function of I and the possible vibrational frequencies of the nucleus. Another region of strong deformation and associated rotational states is found for a few nuclei around $A = 25$.

A further form of collective motion of which nuclei are capable is vibration Such motion probably exists, but it has not yet been identified with certainty

The collective and shell models are complementary and are not mutually exclusive. The nucleons are still thought of as pursuing their effectively free motions inside the deformed and rotating nucleus

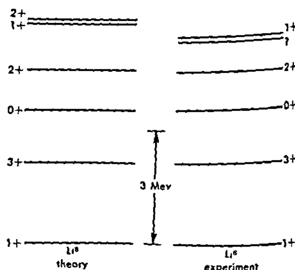
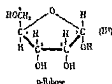
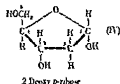


Fig 7. Comparison between the experimental level scheme of Li^6 and that calculated using the shell model in intermediate coupling The symbols on the levels are the spins and parities. The comparison is not continued beyond the point where the experimental identifications become questionable.



D-Ribose

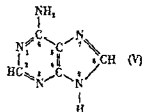


2 Deoxy D-ribose

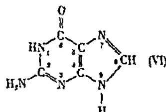
(nucleic acid sugars)

including that of yeast, also contain trace amounts of methylated adenines, for example, 2 methyla-

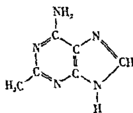
In nucleotides, these subunits are linked together in the sequence: nitrogenous base-ribose-phos-



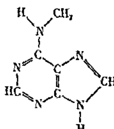
Adenine (6-aminopurine)



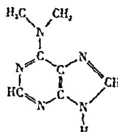
Guanine (2-amino-6-oxypurine)



(VII)
2-Methyladenine

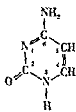


(VIII)
6-Methylaminopurine

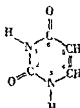


(IX)
6-Dimethylaminopurine

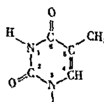
(purines occurring in nucleic acids)



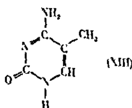
(X)
Cytosine
(2-oxy 6 aminopyrimidine)



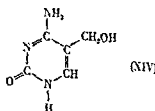
(XI)
Uracil
(2,6-dioxypyrimidine)



(XII)
Thymine
(2,6-dioxy 5 methylpyrimidine)



5-Methylcytosine



5-Hydroxymethylcytosine

(pyrimidines occurring in nucleic acids)

because they would then be *in the same quantum* state. By application of charge independence the interaction with opposed spins is not strong enough to bind them because it did not suffice for a neutron and proton.

Nuclear forces are due to the exchange between nucleons of π-mesons of mass $m_\pi = 273m_e$; this may or may not *interchange either or both the* charges of the particles or their spins. The emission of a π-meson by a nucleon involves an energy of at least $m_\pi c^2$ by the Einstein mass-energy relation (c is the velocity of light). This energy is not really available, but if the meson is reabsorbed within a time τ, then by the Heisenberg uncertainty principle the energy of the system cannot be determined to better than $\hbar/\tau$, and so energy need not be conserved to this degree for this time This suggests that $\tau \sim \hbar/(m_\pi c^2)$. But in this time the meson cannot cover a greater distance than $\tau c \sim \hbar/(m_\pi c) \sim 10^{-13}$ cm. This explains why the nuclear force contains a characteristic distance beyond which it falls rapidly as noted and suggests that in the preceding expression for the potential $r_0 \sim \hbar/(m_\pi c)$. This is consistent with experiment *See* MESON; NUCLEAR REACTION; QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC.

[D H.W.]

Bibliography: R. D. Evans, *The Atomic Nucleus*, 1955; D Halliday, *Introductory Nuclear Physics*, 2d ed., 1955; I. Kaplan, *Nuclear Physics*, 1955.

Nucleation

The formation within an unstable, supersaturated solution of the first particles of precipitate capable of spontaneous growth into large crystals of a more stable solid phase. These first viable particles, called nuclei, may either be formed from solid particles *already present in the system (heterogeneous nucleation)*, or be generated spontaneously by the supersaturated solution itself (homogeneous nucleation).

Heterogeneous nucleation involves the adsorption

may then grow into a large crystal Because the crystal lattice of the foreign solid is in general not the same as that of the solid to be precipitated, the first few layers *are* deposited *in a* lattice configuration which is strained, that is, less stable than the normal lattice of the precipitating material. The degree of lattice strain determines the effectiveness of a given heterogeneous nucleating agent Thus, a material whose crystal structure is greatly different from that of the solid to be precipitated *will* not bring about precipitation unless the solution is fairly highly supersaturated, whereas, if the solution is seeded by adding small crystals of the precipitating substance itself, precipitation can occur at a *concentration* only slightly higher than that of the saturated solution

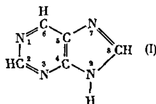
If elaborate precautions are taken to exclude solid particles, it is possible to obtain systems in *which the necessary precipitation nuclei* are spontaneously generated within the supersaturated solution by the process of homogeneous nucleation. In a solution, ions interact with each other to form clusters of various sizes. These clusters in general do not act as *nuclei*, but instead, redissociate into ions. However, if the solution is sufficiently supersaturated *so that its tendency to deplete* itself by deposition of ions onto the clusters overcomes the tendency of the clusters to dissociate, the clusters may act as nuclei and grow into large crystals. The rate at which suitable nuclei are generated within the system is strongly dependent upon the degree of supersaturation. For this reason, solutions which are not too highly supersaturated appear to be stable indefinitely, whereas solutions whose concentration is above some limiting value (the critical supersaturation) precipitate immediately

Nucleation is significant in analytical chemistry because of its influence on the physical characteristics of precipitates Processes occurring during the nucleation period *establish the rate of precipitation* and the number and size of the final crystalline particles. *See* COLLOID; FLOCCULATION; PRECIPITATION (CHEMISTRY); SUPERSATURATION. [D KL.; L G

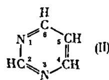
Nucleic acid

The complex nonamino acid *portion of nucleoproteins* containing phosphoric acid, sugars, and purine and pyrimidine bases The two types of nucleic acids are ribonucleic acid (RNA) and deoxyribonucleic acid (DNA)

Ribonucleic acid structure. The ribonucleic acids (RNAs) are long, chainlike, acidic molecules with molecular weights often ranging into the hundreds of thousands, and which sometimes reach 1–2 × 10^6 Their size and complexity is due to the *linking together of large numbers* of simple units called nucleotides The number of chemically different nucleotides involved is relatively small When a typical ribonucleic acid, such as the RNA of yeast, is hydrolyzed in dilute alkali, it yields a mixture of four major nucleotide types (adenylic, guanylic, cytidylic, and uridylic acids) Each of these nucleotide molecules contains three fundamental subunits, (1) a nitrogenous base which is a derivative of either purine (I) or pyrimidine (II) (2) the 5-carbon sugar, D-ribose (III), and (3) phosphoric acid. The different *nucleotides* vary in the nature of their nitrogenous base Some contain the purine derivatives, adenine (V) or guanine (VI). These are the only major purine derivatives found in ribonucleic acids However, certain RNAs

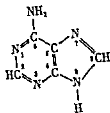


Purine



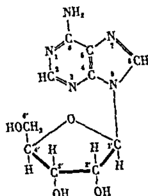
Pyrimidine

Free base



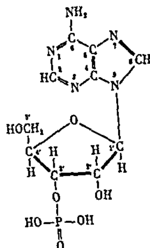
(XV)
Adenine

Nucleoside

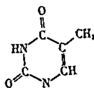


(XVI)
Adenosine

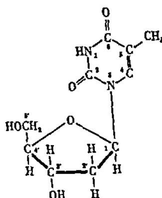
Nucleotide



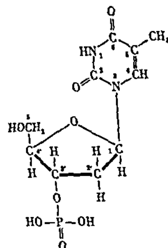
(XVII)
Adenosine-3′-phosphate
(adenylic acid)



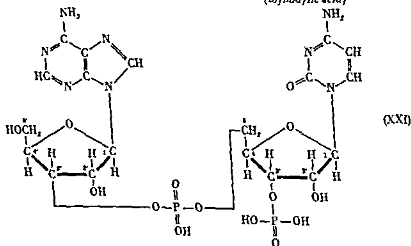
(XVIII)
Thymine



(XIX)
Thymidine



(XX)
Thymidine-3′-phosphate
(thymidylic acid)



(XXI)

Adenylic-cytidylic dinucleotide

phoric acid, (XV) to (XX). Removal of the phosphoric acid leaves the base joined to the sugar; this is called a nucleoside. The ribonucleosides and nucleotides are named for the base they contain, as shown:

Base	*Nucleoside*	*Nucleotide*
Adenine	Adenosine	Adenylic acid
Guanine	Guanosine	Guanylic acid
Cytosine	Cytidine	Cytidylic acid
Uracil	Uridine	Uridylic acid

Molecular formulas of adenine, adenosine, and adenylic acid are shown in (XV), (XVI), and (XVII).

In describing nucleotides, it is necessary to designate the position of the phosphoric acid group on the sugar molecule. In nucleotides derived from RNA by alkaline hydrolysis, the phosphate group is found in the 3′ position. A rearrangement occurs in alkali which also yields the 2′-nucleotides. Yet, when ribonucleic acids are hydrolyzed by specific enzymes, such as pancreatic ribonuclease and a phosphodiesterase prepared from snake venom, the principal products are the nucleoside-5′-phosphates. From degradation studies of this sort, it has been concluded that, in the intact ribonucleic acids, the nucleotides are joined to each other through phosphoryl groups which link the 3′ position of one nucleoside to the 5′ position of the adjacent nucleoside. This type of linkage is illustrated in (XXI) for a dinucleotide of adenylic and cytidylic acids.

In ribonucleic acids, large numbers of nucleotides are joined together in this fashion to form polymers of high molecular weight. Commercial yeast RNA preparations, which are usually somewhat degraded, have molecular weights ranging from 10,000–35,000; certain viral ribonucleic acids have molecular weights exceeding 200,000. The molecular weight of infectious RNA of tobacco mosaic virus has been placed at about 2×10^6. The sequence of the bases in the polynucleotide chain is complex, and no simple repeating pattern of four nucleotides is observed.

Many ribonucleic acids are now prepared by extracting tissue fractions, microorganisms, or isolated viruses with phenol, detergents, or hot salt solutions. In these procedures, the proteins associated with RNA are denatured, and this facilitates their removal from the nucleic acid, which remains in solution. Most ribonucleic acid preparations are heterogeneous, and represent mixtures of different molecular species. The viral RNAs are more uniform in this respect than RNAs prepared from animal tissues.

Ribonucleic acids are usually described chemically by their nucleotide or base composition. For convenience in comparing different data, the amount of adenylic acid is taken equal to 10, and the amounts of the other nucleotides are then expressed in relation to adenylic acid. Some representative analyses are given in Table 1. It can be seen that marked differences in base composition exist between RNAs prepared from different viruses. Evident chemical differences are also found in comparing viral RNAs with those of animal or bacterial cells. There are also chemical and metabolic differences between the ribonucleic acids prepared from different parts of the same cell. Tracer studies have shown that the nucleus is the most active site of cellular RNA synthesis, and that the nucleolus is particularly concerned with the synthetic aspects of RNA metabolism.

Table 1. Nucleotide composition of some ribonucleic acids

Source of RNA	Adenylic acid	Guanylic acid	Cytidylic acid	Uridylic acid
Virus				
Tobacco mosaic virus	10	8.5	6.2	8.8
Tomato bushy stunt virus	10	10	7.4	8.9
Turnip yellow mosaic virus	10	7.6	16.8	9.8
Potato virus X	10	6.2	6.6	6.2
Cucumber virus	10	10	7.5	11.5
Microorganisms				
Escherichia coli	10	10.2	8.5	8.3
Serratia marcescens	10	10.2	8.5	8.3
Bakers' yeast	10	12	8.0	9.8
Animal tissues				
Starfish eggs	10	15	14	11
Chicken liver	10	17.1	13.6	10.6
Rat liver	10	17.6	14.3	10.8
Calf liver	10	17.9	14.9	8.4
Calf spleen	10	19.7	17.7	8.7
Calf thymus	10	17.3	13.5	11.4
Subcellular fractions				
Thymus nuclei	10	15.7	13.5	12.5
Thymus cytoplasmic fraction	10	17.9	14.4	10.3
Rat liver nuclei	10	14.8	14.3	12.9
Rat liver microsomes	10	16.9	14.7	10.3

Other chemical methods used to characterize the ribonucleic acids include determination of the ribose content, and nitrogen and phosphorus analyses. Free RNA contains 15–16% nitrogen and 8.5–9.0% phosphorus. One of the most striking physical properties of the nucleic acids, their high absorption of ultraviolet light of wavelength 260mμ, is widely used for estimating nucleic acid concentration, and for characterizing the different nucleotides and bases.

Ribonucleic acid function. Some of the biological properties of the ribonucleic acids have been discussed in the article on nucleoprotein. The main role of ribonucleic acid appears to be that of mediating the synthesis of cell proteins. Some ribonucleic acids appear to play a direct part in this process, acting as carriers of the amino acids after they are activated. Several theoretical schemes have been proposed to explain how the complex sequence of 20 or more amino acids in a protein can be determined by a changing sequence of only four nucleotides in a polynucleotide chain, but the actual way in which nucleic acids determine the specific structural configura[illegible]n of pr[illegible] not known.

observations fit a molecular model of DNA which has been proposed as a result of x ray studies of DNA structure. The x-ray diffraction patterns of deoxyribonucleic acids indicate that the polynucleotide chain falls into a helical configuration, resembling a coiled spring. The backbone of the DNA, that is, the -deoxyribose-phosphate-deoxyribose- chain, forms the coil of the helix; the bases jut inward toward the axis of the helix. The x-ray evidence suggests that two spiral DNA chains are closely interlocked, with the bases on one chain paired off, presumably by hydrogen bonding, with bases on the adjacent chain. Spatial limitations impose the restriction that purine be paired with pyrimidine. This restriction may constitute the structural basis for the fact that almost all DNAs have purine pyrimidine ratios close to unity.

The problem of DNA heterogeneity has not yet been completely resolved. However, many preparations of isolated DNA can be fractionated by chromatography and by other means to yield DNA fractions of varying base composition. The question of DNA heterogeneity has special interest because of the role of deoxyribonucleic acid as the genetic material (for discussion *see* NUCLEOPROTEIN). The great diversity of genetically determined characters ... [illegible]

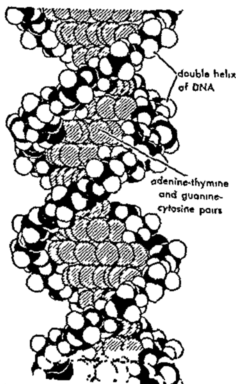
can contain and transmit the *complex hereditary* information which is essential to the life of every organism.

Biosynthesis. The knowledge of the nucleic acids has been enlarged by the discovery by S. Ochoa and A. Kornberg, of enzymes that promote the production of DNA and RNA from smaller organic molecules available commercially. Ochoa's contribution was an enzyme from *Acetobacter vinelandii* that produced RNA from a proper substrate. (*A. vinelandii* is a bacterial species that turns alcohol into acetic acid.) Kornberg's contribution was an enzyme from *Escherichia coli* that performed the same function for DNA. (*E. coli* is a rod-shaped bacterium that is ordinarily found in the intestinal tract.)

Deoxyribonucleic acid biosynthesis. Because DNA is the substance which contains the detailed chemical code governing the heredity of cells, its duplication during cell division requires an extremely precise mechanism of replication and control. The way in which a new DNA molecule arises with the same nucleotides arranged in the same sequence as the [illegible]
most absorb [illegible]
The present [illegible]
nism is large [illegible] of Kornberg, who received the Nobel award in medicine (1959) for his brilliant experiments on the enzymes concerned with DNA biosynthesis.

The synthes[illegible]
of the [illegible]
phosph[illegible]
its num[illegible] example, the phosphoryla-tion of thymidine or thymidylic acid (see XVIII, XIX, XX) gives rise to thymidinetriphosphate, or TTP. The enzymes which phosphorylate nucleotides in this way are called kinases. The nucleoside triphosphates are the substrates for the enzyme (a polymerase) which synthesizes the high polymer, DNA. For DNA synthesis all four deoxynucleoside triphosphates—deoxyadenosinetriphosphate, or dATP; deoxyguanosinetriphosphate, or dGTP; deoxycytidinetriphosphate, dCTP; and thymidinetriphosphate, TTP—*must* be present. DNA polymerases can be prepared *from bacterial* and animal cells which catalyze the *condensation of these four de*oxynucleotides to *form a polynucleotide* with all the properties of DNA. However, *this* synthetic process takes place only in the presence *of a primer* which initiates and controls the synthesis. *The nec*essary primer is deoxyribonucleic acid. This is a *remarkable instance of a molecule's controlling* its own biosynthesis. When small amounts of different DNAs are used as *primers* in this system, the nucleic acid produced resembles the primer in its base *composition* and physical properties. This close correspondence *in structure* between primer and product probably has *its basis* in the pairing of DNA molecules to form a *double helix*. The spatial re*quirements* of this structure are best *met* when adenine is paired opposite thymine and *guanine* is paired with cytosine. This allows the formation of *hydrogen* bonds which act to stabilize the helical [illegible]

Watson Crick model of DNA structure, based orig-



W[illegible] model of DNA [illegible]

Deoxyribonucleic acids structure. Like the ribonucleic acids, the deoxyribonucleic acids (DNAs) are chainlike, polymeric molecules made up by the end-to-end linkage of many nucleotides. The great size and complexity of the DNAs is indicated by their high molecular weights, which range into the millions. The most commonly studied deoxyribonucleic acid, the DNA of the thymus gland, has an average molecular weight of about 6 × 10⁶.

The deoxyribonucleic acids differ chemically in several ways from ribonucleic acids. One of the chief differences, and the one used as the basis for classifying nucleic acids, is in the sugar component. DNAs contain the sugar, 2-deoxy-D-ribose (IV) which differs from ribose in that carbon atom 2 is lacking an attached oxygen atom. Another major difference between DNAs and RNAs is in the nature of their pyrimidine bases. DNAs are lacking in uracil; instead they contain the related base, thymine (XII). Most DNAs contain cytosine (X) as their other major pyrimidine base, but some of the bacteriophages yield a DNA with 5-methylcytosine or 5-hydroxymethylcytosine (XIII, XIV). Many animal and plant DNAs, wheat germ DNA in particular, also contain small amounts of 5-methylcytosine.

All the deoxyribonucleic acids so far investigated contain the two purine bases, adenine and guanine. In the DNA nucleotides, these are joined to the deoxy-sugar by a glycosidic linkage at the carbon atom in position 9 of the ring, as in adenylic acid (XVII). The pyrimidine deoxynucleotides contain deoxyribose linked to position 3 of the base. This is illustrated in (XX) and (XIX) for thymidylic acid and for the corresponding nucleoside, thymidine

In the intact DNA molecule, the deoxynucleotides are joined together through their phosphate groups, which act as bridges between the 3′ position of one nucleoside and the 5′ position of the adjacent nucleoside. The sequence of the different bases in the polynucleotide chain is complex, and no simple repeating pattern is observed.

Unlike RNAs, DNAs are not readily degraded to their component nucleotides by treatment with alkali. However, all the free bases can be obtained by treatment with strong acids at elevated temperatures. This step is usually followed by chromatographic separation of the bases on paper or ion-exchange columns. The bases can be identified and their concentrations determined by their characteristic absorption spectra in the ultraviolet. Some representative DNA analyses are given in Table 2, where the proportions of the different bases are expressed as moles of nitrogenous base per 100 gram atoms of phosphorus.

The DNAs prepared from different animal species differ in their base composition. Thus the chemical composition of DNA appears to be species specific. On the other hand, no significant differences are observed between DNAs prepared from different organs of the same animal Some of the viral DNAs not only have peculiar base compositions, but several bacteriophage DNAs also contain appreciable amounts of the 6-carbon sugar, glucose. This is found only in the phages which contain 5-hydroxymethylcytosine in their DNA; the glucose is linked to this base through the hydroxymethyl group (XIV)

There are certain regularities which appear when the compositions of different DNAs are compared (1) There is generally a 1:1 correspondence between the amounts of purine nucleotides and pyrimidine nucleotides, (2) in many DNAs the molar ratio of guanine to cytosine is equal to 1; and (3) the ratio of adenine to thymine is also 1 These

Table 2. Purine and pyrimidine contents of some deoxyribonucleic acids

Source of DNA	Adenine	Guanine	Cytosine	Thymine	5-Methyl-cytosine	5-Hydroxymethyl-cytosine
Virus						
E. coli bacteriophage T 5	30.3	19.5	19.5	30.8		
E. coli bacteriophage T 2r	32.5	18.2		32.6		16.7
Vaccinia virus	29.5	20.6	20.0	29.9		
Microorganisms						
Pneumococcus type III	29.8	20.5	18.0	31.6		
Yeast	31.3	18.7	17.1	32.9		
Plant						
Wheat germ	*27.3*	*22.7*	*16.8*	*27.1*	*6.0*	
Animal tissues						
Trout sperm	29.8	22.5	20.2	27.5		
Sea urchin sperm (*Arbacia lixula*)	31.2	19.1	19.2	30.5		
Turtle erythrocytes	28.7	22.0	21.3	27.9		
Hen erythrocytes	28.8	20.5	21.5	29.2		
Human thymus	30.9	19.9	19.8	29.4		
Calf thymus	28.0	23.5	20.4	28.1		
Pig thymus	30.0	20.4	20.7	28.9		
Pig thyroid	30.0	20.8	20.7	28.5		
Pig spleen	29.8	20.4	20.8	29.2		

enera died out in the Eocene The only surviving enus is *Apatopygus* (see illustration) from shalow and deep waters off New Zealand and Australia. ee ECHINOIDEA; EUECHINOIDEA [H B F.]

Nucleon

A collective name for a proton or neutron Protons and neutrons are the main constituents of the nuclei of atoms and have considerable similarity between themselves They have the same spin, the same statistics, approximately the same mass, and, through the process of beta decay, can transform into each other. For this reason it is convenient to have a common term to designate them both.

Occasionally neutrons and protons are considered as two states of a single particle called the nucleon, the two states being distinguished by the special value of an internal variable which can then assume only two values and is called the third component of the isotopic spin. *See* ISOTOPIC SPIN, *see also* ELEMENTARY PARTICLE; NEUTRON; PROTON. [E.C S.]

Nucleonics

The technology based on phenomena of the atomic nucleus These phenomena include radioactivity, fission, and fusion (see NUCLEAR PHYSICS) Thus, nucleonics embraces such devices and fields as nuclear reactors, radioisotope applications, radiation-producing machines (such as cyclotrons and Van de Graaff accelerators), the application of radiation for biological sterilization and for the induction of chemical reactions, and radiation detection devices (see NUCLEAR ENGINEERING, PARTICLE ACCELERATOR) Nucleonics makes use of and serves virtually all other technologies and scientific disciplines

That part of the industry concerned with nuclear reactors involves a cross section of the entire industrial complex the chemical industry is concerned with uranium ore refining, fuel and moderator preparation, and fuel reprocessing, the light and heavy metals industry, with fuel fabrication, special component fabrication to withstand environmental conditions including radiation, and containment materials, the machinery industry with control rods, fuel charge and discharge devices and manipulators, and the instrument industry with control systems The many applications of nuclear reactors and isotopes also bring the industries making use of them into the field so that electrical generation, marine and aircraft propulsion, process heat, special industrial devices, and agriculture to name a few are industries participating to some degree in nucleonics.

A number of service activities such as reactor-design consultation, film badge reading, special shipping and disposal of radioactive nuclear materials and wastes, and analytical services by such techniques as low level counting and activation are included in the nucleonics industry The unique radiation hazards and benefits associated with nuclear technology have also engendered special legal, political, and mercantile aspects. [B.I.S.]

Nucleoprotein

A generic term for a great class of conjugated proteins in which large acidic molecules called nucleic acids are linked to or are closely associated with molecules of protein. Nucleoprotein complexes occur in all living animal, plant, and bacterial cells, where they play vital roles in cell duplication and protein synthesis. Simpler forms, such as the animal, plant, and insect viruses and the bacteriophages which infect bacteria, are [illegible]

[illegible]

and nucleic acids in particular, form the chemical basis of heredity, and it is very probable that life as we know it always involves the presence and activity of these complex molecules *See* BACTERIA; BACTERIOPHAGE; VIRUS.

Distribution, type, and function. There is a characteristic and significant distribution of nucleoproteins according to chemical type in nearly all cells. There are two groups of nucleic acids, the deoxyribonucleoproteins [illegible]

[illegible] ... *deoxyribose*.

Deoxyribonucleoproteins. Those complexes in which the nucleic acid component contains the sugar, deoxyribose, are localized in the cell nucleus. Their presence there can be shown by examining cells under the microscope after specific staining reactions for deoxyribonucleic acid (DNA), and also by a chemical analysis of isolated nuclei.

The deoxyribonucleoproteins occur in the cell nucleus as the main components of the structure of the [illegible]

[illegible] ... cells which contain only a single chromosome set Cells in the [illegible]

[illegible] ... mechanisms The direct demonstration of a genetic role of DNA came from experiments on bacterial transformation when it was discovered [illegible] from a ca[illegible] bacteria c[illegible] out such [illegible] recipient [illegible] ties of the [illegible], that is they formed a capsule, or became drug resistant. The change induced by the DNA is inherited and persists through succeeding generations of the transformed bacteria

inally on x-ray diffraction patterns of DNA fibers (see illustration).

When the coiled DNA molecules separate, prior to their replication, each one may act as a primer, controlling the synthesis of new DNA molecules of appropriate base composition. In this way more DNA is made resembling the original primer material. There is now good evidence that the double DNA helix is uncoiled by other enzymes which enable primer action to take place. This may be the function of the small amount of protein which is injected together with bacteriophage DNA when this virus infects a susceptible bacterium.

Attempts are now being made to synthesize enzymatically DNAs with characteristic biological properties, and to use these molecules in transforming the hereditary characteristics of bacterial species.

Ribonucleic acid biosynthesis. The biosynthesis of ribonucleic acids can be considered in two parts. The first is concerned with the sequence of the terminal nucleotides in some RNA chains of relatively low molecular weight. These are the RNAs which are believed to play an important role in protein biosynthesis by serving as amino acid carriers; that is, the molecules transport the individual amino acids to sites where specific proteins are formed. Carrier RNAs occur in tissue homogenates and in some subcellular fractions, including the nuclei. Such homogenates may also contain enzyme systems concerned with the addition of specific nucleotides to the carrier RNA chain. Many studies have indicated that the terminal nucleotide sequence being synthesized is adenylic acid—cytidylic acid—cytidylic acid, with the latter nucleotide attached to the rest of the RNA molecule. This process of adding three nucleotides to an already existing RNA chain requires the presence of the enzymes and of the ribonucleoside triphosphates (ATP and CTP) as precursors. This important facet of RNA synthesis is of general interest because a specific nucleotide sequence is controlled enzymatically, and because the reaction stops when the sequence is complete. These factors may be of value in understanding the more general problem of total RNA synthesis.

Total RNA synthesis is a major problem, and the mechanisms by which many nucleotides are assembled in a precise sequence to form a biologically active RNA molecule are being studied from many angles. One of the great illuminating steps in describing the synthesis of such complex molecules was made by Ochoa (Nobel Laureate in 1959) in his work on the polynucleotide phosphorylases of bacterial cells These enzymes catalyze the synthesis of polynucleotides of high molecular weight. Their substrates are the ribonucleoside diphosphates: adenosinediphosphate (ADP), guanosinediphosphate (GDP), cytidinediphosphate (CDP), and uridinediphosphate (UDP). This RNA-synthesizing system differs in several respects from the process of DNA biosynthesis described previously. (1) The substrates are nucleoside diphosphates, not triphosphates. (2) All four nucleotides need not be present; polymers can be made containing only one or two nucleotide species. (3) Although the reaction can be primed by RNA (and by other polyribonucleotides) the composition of the product does not seem to be determined completely by the nucleotide sequence of the primer. This may indicate the need for other structural determinants in the reaction mixture.

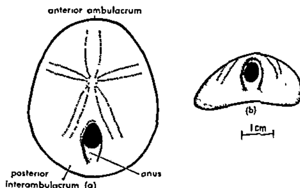
The polynucleotide phosphorylases of bacteria have been extremely useful in producing a variety of synthetic polynucleotides of known structure Their chemical and physical properties are important guides in the study of nucleic acid structure and spatial configuration.

Other approaches to the problem of RNA biosynthesis indicate that mammalian tissues contain enzymes which catalyze the incorporation of nucleoside triphosphates (for example, uridinetriphosphate) into RNA. It remains to be seen whether these reactions are on the primary route of total RNA synthesis in animal tissues. *See* AMINO ACIDS; DEOXYRIBONUCLEIC ACID; ENZYME; RIBONUCLEIC ACID; VIRUS. [V.G.A.]

Bibliography: A. Kornberg, *Enzymatic Synthesis of Deoxyribonucleic Acid*, The Harvey Lectures, ser. 54, 1958–1959; J. S. Roth (ed.), Enzymes of polynucleotide metabolism, *Ann. N.Y. Acad Sci.*, 81(3), 1959.

Nucleolitoida

An order of exocyclic Euechinoidea in which the ambulacra are all similar, usually with the inner and outer pores unequally developed. There are no fascioles, no teeth in the adult, and the genital plates remain separate. Of the two families, the Galeropygidae are round or ovoid forms with simple ambulacra Exclusively Tethyan (from the ancient sea which preceded the present Mediterranean), the family ranged from the Lias of the Early Jurassic to the Cenomanian of the Upper Cretaceous. The Nucleolitidae are similar, but differ in having phyllodes around the mouth, or (as in *Catopygus*) a complete floscelle (*see* CASSIDULOIDA) This family arose in the Jurassic and became cosmopolitan in the Cretaceous, but most



Apatopygus recens, an extant New Zealand nucleolitoid (a) Aboral aspect. (b) Posterior aspect

genera died out in the Eocene The only surviving genus is *Apatopygus* (see illustration) from shallow and deep waters off New Zealand and Australia. See ECHINOIDEA; EUECHINOIDEA. [H.B.F.]

Nucleon

A collective name for a proton or neutron. Protons and neutrons are the main constituents of the nuclei of atoms and have considerable similarity between themselves. They have the same spin, the same statistics, approximately the same mass, and, through the process of beta decay, can transform into each other For this reason it is convenient to have a common term to designate them both

Occasionally neutrons and protons are considered as two states of a single particle called the nucleon, the two states being distinguished by the special value of an internal variable which can then assume only two values and is called the third component of the isotopic spin. See ISOTOPIC SPIN; see also ELEMENTARY PARTICLE, NEUTRON; PROTON. [E.G.S.]

Nucleonics

The technology based on phenomena of the atomic nucleus. These phenomena include radioactivity, fission, and fusion (see NUCLEAR PHYSICS). Thus, nucleonics embraces such devices and fields as nuclear reactors, radioisotope applications, radiation-producing machines (such as cyclotrons and Van de Graaff accelerators), the application of radiation for biological sterilization and for the induction of chemical reactions, and radiation-detection devices (see NUCLEAR ENGINEERING, PARTICLE ACCELERATOR) Nucleonics makes use of and serves virtually all other technologies and scientific disciplines.

That part of the industry concerned with nuclear reactors involves a cross section of the entire industrial complex: the chemical industry is concerned with uranium ore refining, fuel and moderator preparation, and fuel reprocessing. the light and heavy metals industry, with fuel fabrication, special component fabrication to withstand environmental conditions including radiation, and containment materials, the machinery industry with control rods, fuel charge and discharge devices, and manipulators, and the instrument industry with control systems The many applications of nuclear reactors and isotopes also bring the industries making use of them into the field, so that electrical generation, marine and aircraft propulsion, process heat, special industrial devices, and agriculture, to name a few, are industries participating to some degree in nucleonics.

A number of service activities such as reactor-design consultation, film-badge reading, special shipping and disposal of radioactive nuclear materials and wastes, and analytical services by such techniques as low level counting and activation are included in the nucleonics industry The unique radiation hazards and benefits associated with nuclear technology have also engendered special legal, political, and mercantile aspects. [B.I.S.]

Nucleoprotein

A generic term for a great class of conjugated proteins in which large acidic molecules called nucleic acids are linked to or are closely associated with molecules of protein. Nucleoprotein complexes occur in all living animal, plant, and bacterial cells, where they play vital roles in cell duplication and protein synthesis. Simpler forms, such as the animal, plant, and insect viruses and the bacteriophages which infect bacteria, are largely nucleoprotein in composition. In both cells and viruses, there is good experimental evidence that nucleoproteins, and nucleic acids in particular, form the chemical basis of heredity, and it is very probable that life as we know it always involves the presence and activity of these complex molecules. See BACTERIA; BACTERIOPHAGE; VIRUS.

Distribution, type, and function. There is a characteristic and significant distribution of nucleoproteins according to chemical type in nearly all cells. There are two groups of nucleic acids, the deoxyribonucleoproteins and the ribonucleoproteins, which differ in the nature of the pentose or 5-carbon sugar which they contain. The pentose can either be ribose or deoxyribose.

Deoxyribonucleoproteins. Those complexes in which the nucleic acid component contains the sugar, deoxyribose, are localized in the cell nucleus. Their presence there can be shown by examining cells under the microscope after specific staining reactions for deoxyribonucleic acid (DNA), and also by a chemical analysis of isolated nuclei.

The deoxyribonucleoproteins occur in the cell nucleus as the main components of the structure of the chromosomes. An interesting corollary of this chromosomal localization of deoxyribonucleic acid is the [illegible] with two sets of chromosomes, have twice the DNA content of sperm cells which contain only a single chromosome set. Cells in the [illegible] mechanisms The direct demonstration of a genetic role of DNA came from experiments on bacterial transformation, when it was discovered that deoxyribonucleic acid prepared from a capsule-forming or a drug-resistant strain of bacteria could be transferred to another strain without such properties As a result of this transfer, the recipient bacteria took on some of the characteristics of the DNA donor, that is they formed a capsule, or became drug resistant. The change induced by the DNA is inherited and persists through succeeding generations of the transformed bacteria

inally on x-ray diffraction patterns of DNA fibers (see illustration).

When the coiled DNA molecules separate, prior to their replication, each one may act as a primer, controlling the synthesis of new DNA molecules of appropriate base composition. In this way more DNA is made resembling the original primer material. There is now good evidence that the double DNA helix is uncoiled by other enzymes which enable primer action to take place. This may be the function of the small amount of protein which is injected together with bacteriophage DNA when this virus infects a susceptible bacterium.

Attempts are now being made to synthesize enzymatically DNAs with characteristic biological properties, and to use these molecules in transforming the hereditary characteristics of bacterial species.

Ribonucleic acid biosynthesis. The biosynthesis of ribonucleic acids can be considered in two parts. The first is concerned with the sequence of the terminal nucleotides in some RNA chains of relatively low molecular weight. These are the RNAs which are believed to play an important role in protein biosynthesis by serving as amino acid carriers; that is, the molecules transport the individual amino acids to sites where specific proteins are formed. Carrier RNAs occur in tissue homogenates and in some subcellular fractions, including the nuclei. Such homogenates may also contain enzyme systems concerned with the addition of specific nucleotides to the carrier RNA chain. Many studies have indicated that the terminal nucleotide sequence being synthesized is adenylic acid—cytidylic acid—cytidylic acid, with the latter nucleotide attached to the rest of the RNA molecule. This process of adding three nucleotides to an already existing RNA chain requires the presence of the enzymes and of the ribonucleoside triphosphates (ATP and CTP) as precursors. This important facet of RNA synthesis is of general interest because a specific nucleotide sequence is controlled enzymatically, and because the reaction stops when the sequence is complete. These factors may be of value in understanding the more general problem of total RNA synthesis.

Total RNA synthesis is a major problem, and the mechanisms by which many nucleotides are assembled in a precise sequence to form a biologically active RNA molecule are being studied from many angles. One of the great illuminating steps in describing the synthesis of such complex molecules was made by Ochoa (Nobel Laureate in 1959) in his work on the polynucleotide phosphorylases of bacterial cells. These enzymes catalyze the synthesis of polynucleotides of high molecular weight. Their substrates are the ribonucleoside diphosphates: adenosinediphosphate (ADP), guanosinediphosphate (GDP), cytidinediphosphate (CDP), and uridinediphosphate (UDP). This RNA-synthesizing system differs in several respects from the process of DNA biosynthesis described previously. (1) The substrates are nucleoside diphosphates, not triphosphates. (2) All four nucleotides need not be present; polymers can be made containing only one or two nucleotide species. (3) Although the reaction can be primed by RNA (and by other polyribonucleotides) the composition of the product does not seem to be determined completely by the nucleotide sequence of the primer. This may indicate the need for other structural determinants in the reaction mixture.

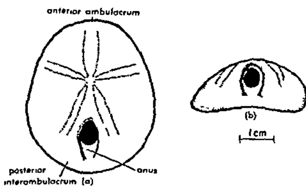
The polynucleotide phosphorylases of bacteria have been extremely useful in producing a variety of synthetic polynucleotides of known structure. Their chemical and physical properties are important guides in the study of nucleic acid structure and spatial configuration.

Other approaches to the problem of RNA biosynthesis indicate that mammalian tissues contain enzymes which catalyze the incorporation of nucleoside triphosphates (for example, uridinetriphosphate) into RNA. It remains to be seen whether these reactions are on the primary route of total RNA synthesis in animal tissues. See AMINO ACIDS; DEOXYRIBONUCLEIC ACID; ENZYME; RIBONUCLEIC ACID; VIRUS [V.G.A.]

Bibliography: A. Kornberg, *Enzymatic Synthesis of Deoxyribonucleic Acid,* The Harvey Lectures, ser. 54, 1958–1959; J. S. Roth (ed.), Enzymes of polynucleotide metabolism, *Ann. N.Y. Acad. Sci.,* 81(3), 1959.

Nucleolitoida

An order of exocyclic Euechinoidea in which the ambulacra are all similar, usually with the inner and outer pores unequally developed. There are no fascioles, no teeth in the adult, and the genital plates remain separate. Of the two families, the Galeropygidae are round or ovoid forms with simple ambulacra. Exclusively Tethyan (from the ancient sea which preceded the present Mediterranean), the family ranged from the Lias of the Early Jurassic to the Cenomanian of the Upper Cretaceous. The Nucleolitidae are similar, but differ in having phyllodes around the mouth, or (as in *Catopygus*) a complete floscelle (see CASSIDULOIDA). This family arose in the Jurassic and became cosmopolitan in the Cretaceous, but most



Apatopygus recens, an extant New Zealand nucleolitoid. (a) Aboral aspect. (b) Posterior aspect

*Amino acid composition of some proteins associated with deoxyribonucleic acids**

Amino acid	Protamines of			Histones of				
						Calf thymus		
	Salmon sperm (salmine)	Rainbow-trout sperm	Fowl sperm (gallin)	Calf liver	Calf kidney	Total histones	Lysine-rich histone	Arginine-rich histone
Arginine	89.6	86.96	76.4	20.2	21.1	21.5	5.3	28.2
Histidine			1.53	3.7	3.5	3.9	0	4.4
Lysine			0	16.4	16.7	19.1	42.1	16.0
Aspartic acid			0.28	3.7	3.5	3.1	1.5	3.4
Glutamic acid			0.45	5.6	5.6	5.3	2.4	6.0
Glycine	1.81		2.80	5.4	5.9	5.4	5.0	6.3
Alanine	0.57	2.08	1.22	8.2	8.5	8.7	19.2	7.7
Valine	1.23	3.50	0.68	3.8	4.0	4.0	3.6	4.7
Leucine	0.57		0.35	5.4	5.3	5.0	3.1	5.8
Isoleucine			0.35	2.9	2.9	2.8	0.67	3.3
Serine	3.94	2.02	3.89	3.7	3.8	3.8	4.9	3.5
Threonine			0.72	3.7	3.8	3.7	4.2	3.9
Cystine			0	0.24	0.19	0.06	0	0.21
Methionine			0	0.81	0.66	0.63	0	0.68
Proline	2.30	6.11	1.94	2.8	3.3	3.2	7.1	2.7
Phenylalanine			0	1.3	1.3	1.4	0.39	1.2
Tyrosine			1.58	1.6	1.7	1.6	0.37	1.9
Tryptophan			0	0	0	0	0	0
Amide-NH_3			0.17	5.1	4.5	4.1		

* Results are expressed as percentage of the total protein nitrogen contributed by each of the amino acids.

protamine are easily isolated from fish sperm nuclei by extraction in concentrated salt solutions. Free protamines can be extracted from nucleoprotamine preparations or from sperm nuclei with dilute acid under conditions where the DNA remains insoluble.

Histones. The nuclei of somatic or body cells and of certain types of sperm contain other basic proteins known as histones. These have a higher molecular weight than do the protamines, as a consequence, histones, unlike protamines, will not dialyze through a cellophane membrane. They also have a more [illegible]

[illegible] proteins of the nucleus, they contain no tryptophan. *See* AMINO ACIDS.

As basic proteins, histones are capable of saltlike combination with deoxyribonucleic acid. Such electrostatic linkages involving DNA, which is localized in chromosomes, are probably important to chromosome structure and function, but there is good evidence that the bonds which hold histones in chromosomes are not entirely limited to salt-like [illegible]

Histone-DNA complexes can be obtained from the nuclei of many tissues and from isolated chromosomes. Such complexes have characteristic solubility properties which facilitate their isolation. They are soluble in water and in very strong salt solutions, such as 1 *M* NaCl, but they are insoluble at intermediate salt concentrations. For this reason, the isolation of nucleohistones usually involves an extraction in strong salt solutions. This gives a viscous solution of nucleohistone, which can then be precipitated by lowering the salt concentration of the extract. The histones themselves are usually prepared by extracting nuclei or nucleohistone preparations with dilute acid. [illegible]

[illegible] geneous. Several different types of histones exist in an acid extract of nuclei. Experimentally, a clear distinction can be made between those histones which precipitate from alkaline solution, at pH 10–11, and some basic proteins which are not precipitable in this way. The latter proteins can also be separated by other types of fractional precipitation and by chromatography. They are characterized by a very high lysine content (see table). A purified, lysine-rich histone prepared from calf thymus is reported to have a molecular weight of 18,000.

Some of the histones prepared from different organs of the same animal resemble each other very closely in their [illegible]

[illegible] known it has been suggested that these positive proteins may serve to mask the negative charges of the DNA molecule. Because this negative electrical charge seems to be correlated with the chemical activity of the nucleus, histones would act to retard or limit nuclear activity.

In some viruses, notably the bacteriophages, the mechanism of infection also suggested that DNA is the hereditary material. The virus attaches to the bacterial cell wall and injects its DNA into the cell, leaving its protein capsule behind. Once inside the host cell, the viral DNA sets into operation the mechanisms which result in a new generation of virus particles.

Ribonucleoproteins. The second great group of *nucleoproteins, the ribonucleoproteins,* is characterized by a nucleic acid component which contains ribose. The ribonucleoproteins occur in some viruses, in bacteria, and in both the nucleus and the cytoplasm of animal and plant cells. In the nucleus, much of the ribonucleoprotein is localized in the dense structure called the nucleolus, but some is also distributed along the length of the chromosomes. In the cytoplasm of many cells, the ribonucleoproteins are found in the complex network of membranes and granules called the ergastoplasm or endoplasmic reticulum. Because of their content of ribonucleic acid (RNA), these structures are basophilic, that is they stain with basic dyes Because all nucleic acids absorb light strongly in the ultraviolet region of the spectrum, cell structures which are rich in RNA also appear dark under the ultraviolet microscope. When different types of cells are examined for their ribonucleic acid content, the results show a striking parallelism between the amount of RNA present in a cell and its capacity to synthesize proteins. Thus, the cells of the pancreas, which must secrete large amounts of digestive enzymes, may contain over 20% of their mass as RNA, while red cells which no longer synthesize the protein, hemoglobin, have little or no ribonucleic acid. This sort of parallelism suggested a role of ribonucleoprotein in the process of protein synthesis More direct evidence came from studies of protein synthesis in broken bacterial cells and in cell fractions rich in RNA such as the microsome fraction. This fraction can be prepared from homogenized animal tissues. Extraction of the microsomes with a detergent (deoxycholate) disperses the membranes and leaves the small granules These contain up to 50% of their mass as RNA. They are believed to be the primary site of protein synthesis in the cytoplasm. It was found that when a specific enzyme, ribonuclease, attacks RNA the uptake of amino acids into the proteins in these systems is then stopped. Recent work on proteins of the microsome fraction shows that the proteins bound to RNA are rich in the basic amino acids arginine and lysine, but unlike the histones of the cell nucleus, they contain tryptophan. As yet, little is known about their fractionation and chemical composition. However many different cytoplasmic proteins also occur in association with the microsome fraction. Further experiments indicated that some RNAs can act as a carrier of amino acids after they have "activated." These amino acids are then linked to form different proteins on or in cell structures which are rich in ribonucleoprotein.

Other evidence which shows that ribonucleic acid can determine the specific nature of a protein stems from work on plant viruses, where RNA itself can function as the infectious agent. When the RNA prepared from one strain of tobacco-mosaic virus is combined with the proteins prepared from other strains of the virus and used to infect a plant, the progeny of these hybrid viruses contain a protein whose properties are determined by the nucleic *acid used, and not by the protein. Thus,* in the tobacco-mosaic virus and in some other plant and animal viruses, ribonucleic acid is a genetic determinant, as DNA is in the bacteriophages, in bacteria and in higher organisms.

Protein components. In nature, the different types of nucleic acid, each of which includes a great variety of species-specific molecules, are joined to a complex array of different proteins The simplest of these nucleic acid-associated proteins comprise the group of very basic, small proteins called protamines, associated with the sperm cells. Histones, another basic small protein, are found in the nuclei of somatic cells and of some sperm cells.

Protamines. The protamines were first discovered in the sperm cells of fish, where they occur in high concentrations in the nucleus, together with deoxyribonucleic acid. They are also found in the mature sperm of other species, such as fowl, but the nuclei of somatic cells contain more complex basic proteins called histones. The fact that chromosome composition can vary is shown in the course of spermatogenesis in fish, where the basic proteins of the nucleus change in type, beginning as histones in the early stages of maturation and ending as the comparatively simple protamines in the mature sperm cell.

The basicity of the protamines is due to their high content of basic amino acids, arginine in particular. In salmine, the protamine of salmon sperm arginine is the only basic amino acid present, but it alone accounts for 89–90% of the total protein nitrogen. The composition of a few protamines is given in the table, and it can be seen that these relatively simple proteins are lacking in tryptophan and the sulfur-containing amino acids. In some protamines, the cyclic amino acid, proline, occurs at the end of the polypeptide chain.

The usual protamine preparations are mixtures of related polypeptides which reveal their heterogeneity when they are examined chromatographically, or separated electrophoretically in an electric field Estimates of the molecular weights of different protamine preparations range from 2000 to 8000, the small size of the protamines permits them to diffuse through a cellophane membrane; this facilitates their separation from larger molecules of nucleic acid and proteins, which cannot pass the membrane.

The basicity of the protamines confers upon them a net positive charge which allows a salt-like combination with the negatively charged DNA of the sperm nucleus. Complexes of nucleic acid and

Sea cucumbers are hermaphroditic and reciprocally fertilizing They creep about on seaweed and other plants and on animals of the sea floor. Most of them feed on seaweed, but some are carnivorous, eating sponges and other animals. Some species feed extensively upon the *hydroids* of various Coelenterata. They seem to be able to eat these hydroids without causing discharge of their nematocysts

Sea cucumbers are sometimes incorrectly called sea slugs *See* MOLLUSCA; MUSSEL; SLUG (ZOOLOGY); SNAIL. [J.D.B]

Nudibranchia

A suborder in the Opisthobranchia containing the sea slugs The shell is usually absent, the gills are variable in size and position, and the mantle cavity is generally lacking. Digestive diverticula branch into a series of tubules, the club-shaped cerata on the dorsal surface of the body

These animals are highly colored and are among the most beautiful animals in the sea They are predatory, feeding mainly on coelenterates They occur in all seas, in shallow water to moderate depths, but reach their greatest diversity in the tropics. Certain forms are pelagic, existing in the open sea *See* OPISTHOBRANCHIA [W J C]

Number indicator tube

Any electron tube capable of visually displaying numerical figures In many electronic circuits and equipments it is desirable to indicate numbers Such a tube can be one of a set which will display, for instance, the magnitude of voltage in digital form This is analogous to electrical clocks which display the time in numerical form

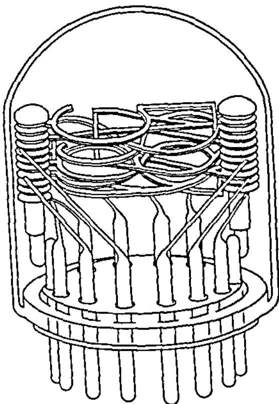
Gas indicator tubes. One of the simplest of such tubes, which is extensively used, is a cold-cathode gas tube in which there is a series of cathodes which light up because of the cathode glow that surrounds any cathode in a gas discharge A common anode is used These cathodes are shaped to correspond to the different numerical digits from 0 through 9 The desired cathode is selected by any suitable switching scheme Cathodes are made of bent wire and are insulated and stacked one above the other so that any particular number can be read when the corresponding cathode is illuminated by applying a suitable voltage to it. *See* ELECTRICAL CONDUCTION IN GASES, GAS TUBE

tir

sc

though there are frequently disturbing reflections on the glass envelope. These can be overcome by putting a polaroid screen, which eliminates reflections, in front of the tubes

Charactron tube. Other types of number-indicating tubes are generally *more* complex in character There is, for instance, a special type of cathode ray tube known as the *Charactron* This



Number indicator tube showing the positions of the number-shaped cathodes. (Burroughs Corp)

tube utilizes an electron beam as in the ordinary cathode-ray tube, which is deflected to pass through a mask which contains different numbers or letters punched on it. The beam is subsequently deflected and focused back to any position on a fluorescent screen so that the particular letter or number through which the beam was passed in the mask appears illuminated on the fluorescent screen Such a tube can be used to display words and sentences and other combinations of numbers and letters Such tubes are, however, rather expensive and require fairly complex circuitry for operation. *See* STORAGE TUBE.

Other devices. Other display devices are available to supplement the available tubes One form of display device for numbers makes use of electro-

illuminating proper combinations of these seven segments, the ten digits, including zero, can be formed Still another number-display device makes use of a stack of transparent plastic strips with the individual digits embossed on them, one on each strip The number on any particular strip can be illuminated by a small light located at the edge of that strip

The development of the above devices and others of a similar character has led to an extensive use of digital display of quantities such as voltage, time, and count *See* COUNTER, DIGITAL. [K.B S]

Other protein fractions of the nucleus. When the DNA and histone are removed from preparations of isolated chromosomes, prepared by disrupting cell nuclei, a residual protein fraction remains. This material often appears as a microfibril which is insoluble in water or neutral salt solutions. It contains a significant amount of ribonucleic acid, the amount depending upon the tissue of origin. The protein component contains tryptophan, a fact which sets it apart chemically from the histones.

The morphological configuration of chromosomes, as seen under the microscope, is dependent on the combination of DNA with the residual proteins. The quantity of residual protein in the chromosomes varies in different tissues; a rough correlation exists between this amount and the size and metabolic activity of the cell. The synthesis of these chromosomal proteins depends on the presence of DNA, but the rate of synthesis varies with changes in the over-all activity of the cell.

Some of the protein of the residual chromosome fraction can be extracted in dilute alkaline solutions and precipitates when the extract is acidified Such protein fractions contain about 10% of firmly bound lipids, which give positive tests for phospholipides and cholesterol.

In addition to histones and residual proteins, the cell nucleus contains many proteins which may or may not be associated with DNA. Some of these are readily extractable in dilute salt solutions. One of the major components of the saline extracts of liver and thymus nuclei is a protein with the properties of a globulin.

Other DNA-associated proteins. Avidin, a protein prepared from egg white, contains some attached deoxyribonucleic acid. This is one of the few instances of DNA occurring outside of the nucleus in animal cells.

Deoxyribonucleic acid occurs together with protein in the bacteriophages and in some insect viruses The bulk of the protein in the bacteriophages forms an outer capsule surrounding the viral DNA. There are three to five different proteins in this capsule; all can be readily separated from the DNA In addition, a small amount of protein occurs in close association with the nucleic acid and is transferred with DNA into the bacterial cell at the time of virus infection.

Many of the insect viruses are considered deoxyribonucleoproteins. In the virus infections of certain insects, such as silkworms, intranuclear inclusion bodies appear. These polyhedral bodies are made up of virus particles embedded in a crystalline protein. The bulk of the polyhedra consists of protein with a molecular weight of about 300,000 The virus itself is a small part of the inclusion and is made up of DNA and a closely associated protein fraction. *See* NUCLEIC ACID

[V.G.A.]

Bibliography: J. Brachet, *Biochemical Cytology*, 1957; E. Chargaff and J. N. Davidson (eds.), *The Nucleic Acids, Chemistry and Biology*, 2 vols., 1955; S. E. Luria, *General Virology*, 1953, W D McElroy and B. Glass (eds.), *A Symposium on the Chemical Basis of Heredity*, 1957.

Nucleus, atomic

Atoms are composed of negatively charged electrons, positively charged protons, and electrically neutral neutrons. The protons and neutrons (collectively known as nucleons) are located in a small central region known as the nucleus The electrons move in orbits which are large in comparison with the dimensions of the nucleus itself. Protons and neutrons possess approximately equal masses, each roughly 1840 times that of an electron. The number of nucleons in a nucleus is given by the mass number A and the number of protons by the atomic number Z. Nuclear radii r are given approximately by $r = 1.4 \times 10^{-13} A^{1/3}$ cm *See* NUCLEAR STRUCTURE.

[H.E.B.]

Nuclide

A species of atom that is characterized by the constitution of its nucleus, in particular by its atomic number Z and its neutron number $A - Z$, where A is the mass number. Whereas the terms isotope, isotone, and isobar refer to *families* of atomic species possessing common atomic number, neutron number, and mass number, respectively, the term nuclide refers to a *particular* atomic species The total number of stable nuclides is approximately 275 About a dozen radioactive nuclides are found in nature, and in addition, hundreds of others have been created artificially

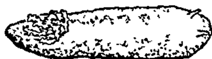
[H.E.B.]

Nuda

A class of the phylum Ctenophora containing the single order Beroida. Their lack of tentacles distinguishes these animals from the class Tentaculata *See* BEROIDA, CTENOPHORA; TENTACULATA

Nudibranch

Any of several marine gastropods which lack, or apparently lack, a shell. They belong to the order Nudibranchia, class Gastropoda, phylum Mollusca They lack true gills and respire most commonly through the skin or by means of secondary gills around the anus, around the edge of the mantle, or in dorsal rows. Many have developed a variety of papillae on their surface which facilitate breathing Some are very elaborate with brilliantly colored papillae and tentacles. Others are slender, smooth-skinned types like the land slugs in appearance Their basic anatomy is similar to that of the mussel and the snail



Nudibranch, *Doris* (From T. I. Storer and R. L. Usinger, *General Zoology*, 3d ed., McGraw-Hill, 1957)

normalization and reassembly of the answer as a "floating word" The system has the advantage of greater control over numbers of widely varying orders of magnitude The disadvantages include slower operation, often by a factor of 5 or more, and occasional unpredictable loss of information.

Besides the operation of complementation, there are other nonarithmetic operations with decimal numbers, for example, comparison Two numbers may be compared for size by a simple inspection of their corresponding digits, beginning from the left and stopping at the first case of inequality. This simple but important property is worth mentioning because comparison is almost impossible in certain other systems. An unusual use of decimal digits is the so-called middle-of-the-square method of generating random numbers. By this method the next 10-digit random number is obtained from the preceding one by squaring the latter and selecting from the square the central 10 digits.

There are many interesting properties of the digits of integer numbers The simpler ones depend on the theory of congruences. The most familiar fact of this sort [illegible]
eve [illegible]
sta [illegible]
If a number is diminished [illegible]
the [illegible]
for [illegible]
out [illegible] known to every school boy. Elevens may be cast out in like manner if the digits are added with alternating signs. Thus, $34162 - (2 - 6 + 1 - 4 + 3) = 34166$, is a multiple of 11 Similarly, grouping the digits by threes, $34535599 - (599 - 535 + 34) = 34535501$ is a multiple of $1001 = 7 \cdot 11 \cdot 13$. This fact is sometimes used to check desk calculator computations by casting out 1001s It is also used to decide quickly whether a given number is divisible by 7, 11, or 13 The number 34535599 is not divisible by 7 11 or 13 since $599 - 535 + 34 = 108$ is not

Squares of integers have digital properties For example, the final digit of a square is either 0, 1, 4, 5, 6, or 9, never 2, 3, 7, or 8. There are only 22 combinations of two digits in which a square can end etc Such facts are sometimes used in finding the factors of a given number by expressing it as a difference of two squares. The rapid recognition of nonsquares is also helpful in many other diophantine problems

The representation of real numbers requires infinite, that is unending decimals If the digits of such a decimal ultimately become periodic the decimal is the ratio P/Q of two integers, and conversely The length of the period is a complicated function of Q, depending on the prime factors of numbers of the form $10^n - 1$ If, and only if Q is of the form $2^a 5^b$, the decimal expansion of P/Q terminates In such cases P/Q has in reality two expansions Thus $7/5 = 1.4000 \ldots = 1.3999 \ldots$

The great majority of decimals do not become periodic, or in other words, almost all real numbers are irrational The class of irrational algebraic numbers, such as the square root of 2, that are roots of polynomials with integer coefficients, is almost completely obscured by other real numbers in their decimal representation. There are only a few statements that can be made about the digits of such numbers other than the obvious one of nonperiodicity. For example, if k consecutive zeros occur, then they cannot occur "too soon" for an infinity of k. On the other hand, almost all real numbers have perfectly normal decimal expansions in the sense that each digit occurs, on the average, one-tenth of the time, each ordered pair one-hundredth of the time, etc. Whether π, e, or $\sqrt{2}$ are normal is not known. The totality of all known examples of normal numbers is countable.

Almost everything that has been said so far about the decimal system applies with equal force and very little modification to a general system based on an integer $b > 1$ instead of 10 The fact that people "know" all the powers of 10 but not the powers of 7 or 12 is purely psychological and based on tradition. Beyond the fact that 10 is even, there is little to recommend it as a base The Babylonians used 60, a large but useful base that is still in vogue for measurement of time and angles. The mathematician J d'Alembert and many others after him, urged the adoption of $b = 12$ with its six divisors. The advent of electronic computers has made a good case for $b = 8$ or some other power of 2. Probably base 8 is used by humans more than any base except 10. For $b > 10$, new characters are needed to represent the extra digits Although there is no agreement as to which characters to adopt, the modern tendency is to use roman letters because they are easily available on the typewriter. The adoption of a second system brings up the question of translating or converting numbers from one system to the other. Methods for doing this are explained in following sections on binary and octal systems.

Binary system. In the binary system every positive integer is the sum of distinct powers of 2 in just one way. Thus $434 = 2^8 + 2^7 + 2^5 + 2^4 + 2^1$, and this is expressed by writing 110110010. The digits corresponding to 2^0, 2^2, 2^3, and 2^6 are zero, since these powers do not occur in 434. The first dozen integers are written as follows:

1	1	4	100	7	111	10	1010
2	10	5	101	8	1000	11	1011
3	11	6	110	9	1001	12	1100

The great advantage of the binary system lies in the fact that there are only two kinds of binary digits, or "bits," namely 0 and 1 This not only gives a simplified arithmetic but provides a language in which to treat two-valued functions or bistable systems Among its disadvantages is the fact that the binary system requires [illegible]
times as [illegible]
d [illegible]

Number systems

Integral numbers may be represented as linear combinations of powers of any convenient and arbitrarily chosen base. The choice of the base is not always made on a rational basis, and number systems have been based on 5, 6, 10, and 60. More recently, systems based on 2 and 8 have proven quite useful in computer applications. The duodecimal number system, in which numbers are represented as linear combinations of powers of 12, has certain advantages because 12 has the factors 1, 2, 3, 4, 6, and 12.

Decimal system. Every positive integer is uniquely a polynomial in 10 with coefficients, called digits, taken from 0, 1, . . . , 9. The fact that

$$205714 = 4 + 1 \cdot 10 + 7 \cdot 10^2 + 5 \cdot 10^3 + 0 \cdot 10^4 + 2 \cdot 10^5$$

is nearly always lost in present-day teaching, and in the hurried application of ordinary arithmetic. In fact, numbers are likely to be thought of as merely an orderly arrangement of their decimal digits.

The decimal method of representing numbers comes from India and Arabia and is only a few centuries old in Europe. The base, 10, is due to the biological fact that man has that many articulate fingers and thumbs. The positional significance, including the meaning and usefulness of zero, is of oriental origin.

The operations of addition and multiplication consist of the corresponding operations with polynomials, together with rules that serve to keep the results inside the system so that they can be used in future operations In the case of addition of two numbers, use is made of either the familiar "carry" rule or the addition table, while for multiplication, use is made of the multiplication table to help represent the product of two digits as a two-digit number. These apparently nonalgebraic operations are so dominant that the basic polynomial structure of the numbers is obscured. Thus, the multiplication of polynomials is done by a more intelligent method than that used for numbers. For example, the multiplication of 2057 by 3416 can be carried out as follows:

```
      2 0 5 7
      3 4 1 6
  -------------
  6,1 7,3 3,4 2
    8,5 3,3 7
  -------------
  7 0 2 6 7 1 2
```

Commas separate those pairs of digits that arise from sums of products of pairs of digits taken one each from the original numbers, and having equal significance. Thus 53, the fourth most significant contribution, is given by

$$53 = 2 \cdot 6 + 0 \cdot 1 + 4 \cdot 5 + 3 \cdot 7$$

This process can be carried out either from right to left or from left to right, and in the latter case, may be terminated when half done if the least significant half of the product is not needed.

In connection with the design and use of automatic computers, in which numbers of a limited size only may be added and multiplied at one time, precautions against overflow in addition, and approximation by rounding in multiplication further complicate the execution of ordinary arithmetic. This creates a system that, strictly speaking fails to satisfy the axioms of arithmetic. This causes serious difficulties in some problems involving millions of additions and multiplications.

Subtraction introduces negative numbers that may be handled by introducing a special digit called a sign digit with its own rules of combination, or by introducing complementation in which the digits of a number are subtracted from 9, except for the last nonzero digit which is subtracted from 10 ... 20570 from 34162 20570 may be c... obtain the

$$
\begin{array}{r}
\ldots 99979430 \\
\ldots 00034162 \\
\hline
\ldots 00013592
\end{array}
$$

Numbers that begin with a run of nines are considered negative in this system Of course care must be taken to guard against overflow in which a very large positive number might be confused with a very small negative one.

Division is a process that can be carried out only rarely with absolute exactness in the decimal system, the process usually being nonterminating This introduces the notion of infinite decimal expansions and the more or less theoretical operations with such numbers In practice, truncation and rounding are used, as in ⅔ = .66667, with consequent errors and departure from the axioms of arithmetic. In this case a quantity like ab/c is not unique but may depend upon the order in which the indicated operations are performed. For complicated and extensive problems involving only the four rational operations of arithmetic, an adequate analysis of the errors involved may be very costly indeed.

Automatic calculation in this simulated real number system may be facilitated by the use of a normalizing coding device called "floating arithmetic." In this system a positive real number is expressed as a truncated decimal between .1 and 1 times the appropriate power of 10 Thus the number π on a 10 digit decimal machine, could be coded 3141592751. In interpreting this "word," the machine separates the last two digits, 51, and subtracts 50 to get the exponent (possibly negative) of the power of 10 by which the mantissa .3141592 would have to be multiplied to obtain π correct to 8 decimals. Rules for multiplying and adding in this system are easily formulated They involve inspection, comparison, and manipulation of the exponents, followed by appropriate shifting right or left of the mantissas, followed next by ordinary decimal arithmetic on the mantissas, and ...

because, in forming the first column, one is, in effect, converting 146 to the binary system.

Another operation in which binary representation is effective is that of raising a given base, B, to a high integer power Suppose that

$$n = b_k b_{k-1} \cdots b_2 b_1 b_0$$

is the binary representation of the integer n. To *compute* B^n most efficiently, form recursively the numbers, w_i, defined by

$$w_0 = B^{b_k} = B$$
$$w_1 = B^{b_{k-1}} (w_0)^2$$

$$w_i = B^{b_{k-i}} (w_{i-1})^2$$

then $w_k = B^n$. In fact

$$w_k = B^{b_0} (w_{k-1})^2 = B^{b_0 + 2(b_1 + 2(b_2 + \cdots))}$$

so that the exponent is

$$b_0 + 2b_1 + 2^2 b_2 + \qquad + 2^k b_k = n$$

Octal system. To write a number in the octal system, once it has been expressed in the binary system, one merely groups the binary digits by threes, beginning at the binary point and working to the left and right. Thus the decimal number 43 4294 gives

(101)(011) (011)(011)(011)(110)(110)(10)

or simply 53.333664 where the last digit should perhaps be 5 On the other hand decimal to octal conversion can be accomplished directly by either of the two methods that correspond in an obvious way to those given for decimal to binary conversion Thus, by subtracting appropriate multiples of powers of 8, beginning with the largest possible power,

$$5280 = 1 \quad 8^4 + 2 \quad 8^3 + 2 \quad 8^2 + 4 \quad 8$$

so that in the octal system th... = 12240 ...

[illegible] and 10 Dividing 10 by 8 gives 2 and 1 This gives the digits in reverse order

Octal to decimal conversion may be effected by the use of a convenient table of powers of 8, a sample of which follows

n	8^n	n	8^n
0	1	−1	125000
1	8	−2	015625
2	64	−3	001953
3	512	−4	000244
4	4096	−5	000031
5	32768	−6	000004

The octal system with its eight digits 0, 1, . . . , 7 affords a convenient way of conden... ter displa... octal syst... metic. Th... as shown

Addition

	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7	10
2	2	3	4	5	6	7	10	11
3	3	4	5	6	7	10	11	12
4	4	5	6	7	10	11	12	13
5	5	6	7	10	11	12	13	14
6	6	7	10	11	12	13	14	15
7	7	10	11	12	13	14	15	16

Multiplication

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	10	12	14	16
3	0	3	6	11	14	17	22	25
4	0	4	10	14	20	24	30	34
5	0	5	12	17	24	31	36	43
6	0	6	14	22	30	36	44	52
7	0	7	16	25	34	43	52	61

Examples of addition and multiplication in the octal system are:

```
 4375          5734
 3704            16
-----         -----
10301         43150
              5734
             ------
             123010
```

Octal arithmetic can be checked by "casting out sevens" (instead of nines) by adding the digits. Thus for the addition problem above

$$4375 \equiv 4 + 3 + 7 + 5 = 23 \equiv 2 + 3 = 5 \pmod 7$$
$$3704 \equiv 3 + 7 + 0 + 4 = 16 \equiv 1 + 6 \equiv 0 \pmod 7$$
$$10301 \equiv 1 + 3 + 1 = 5 \pmod 7$$

Checking by casting out nines involves taking the octal digits with alternating signs. Thus

$$4375 \equiv 5 - 7 + 3 - 4 = -3 \equiv 6 \pmod 9$$
$$3704 \equiv 4 - 0 + 7 - 3 = 8 \pmod 9$$
$$10301 \equiv 1 + 3 + 1 = 5 \equiv 6 + 8 \pmod 9$$

The octal system requires only 10% more digits than the decimal system to represent the same *amount* of information. Some computing systems use base 16, in which case binary information is handled in sets of four bits. This system is more compact than the decimal system, 100 hexodecimals being equivalent to 120 decimals, but it requires a multiplication table with nearly three times as many entries.

Computing systems of binary type have subroutines for the conversion of any kind of decimal information into binary information during *input*, and vice versa during output, so that a facility in octal arithmetic is needed only rarely during checking and testing of a new problem. See DIGITAL COMPUTER, NUMERICAL ANALYSIS. [D.H L.]

Bibliography: T. Dantzig, *Number, the Language of Science*, 1956; C. Reid, *From Zero to Infinity*, 1955.

imal digits into binary form, while the purely binary machines use full binary arithmetic.

The physical representation of binary numbers, or information, is possible in many forms. A row of lights, some on and some off, may be interpreted as a binary number. A set of condensers some charged and some not, or a set of high and low voltages, or a set of magnets with fluxes in one direction or another are electronic examples of media for the processing and retention of data in the binary system. The fact that there are only two states to recognize accounts for the great reliability of such computing systems.

The conversion of decimal, or base 10, integers into the binary system can be done in two ways. First, one may subtract from the given integer the highest power of 2 not exceeding this number and record a 1 in the binary position corresponding to this power of 2. The remainder of this subtraction, if not zero, now replaces the original number and the process is repeated until a zero remainder is obtained

Alternatively, one may divide the given number by 2 and record the remainder, 0 or 1, as the final binary digit. The quotient in this division now replaces the given number and the process is repeated and continued until a quotient of zero is reached. The two methods are illustrated in the case of converting 434 to the binary system:

434		434	
−256		217	0
178	1	108	10
−128		54	010
50	11	27	0010
−32		13	10010
18	1101	6	110010
−16		3	0110010
2	11011	1	10110010
−2		0	110110010
0	110110010		

Both processes have obvious inverses for going from the binary system to the decimal system. In the first case, the indicated powers of 2 are simply added together, and in the second case, a sequence of doubling operations is used, followed by the addition of 0 or 1 as specified by the given binary number. For numbers between 0 and 1 similar procedures are available. Either the subtraction of powers of 2 (negative powers) can be continued or the given number can be doubled, followed by subtraction of whichever of the numbers 0 or 1 will make the remainder lie between 0 and 1, and the operation continued with the remainder as before. The reader may wish to test his understanding by verifying that 43.4294 has the binary representation:

$$101011.01101101111011010\ldots$$

Arithmetic in the binary system is remarkably simple. For addition, only $1 + 1 = 10$ is needed, while the multiplication table reduces to $1 \cdot 1 = 1$. Examples of addition and multiplication are

110101	(53)	1101	(13)
11001	(25)	1011	(11)
1001110	(78)	1101	
		1101	
		1101	
		10001111	(143)

Such simple operations are readily performed electronically with extreme rapidity and reliability

The binary system is useful not only to represent numbers but also to record and process information. In fact the unit of information is a binary digit. For example, given a set, S, of objects and a property P, it is possible to record which objects have the property P, and which do not, by assigning a binary position to each object of S and recording there a 1 or 0 according as the property P is, or is not, possessed by the corresponding object. Thus if the objects are the first odd numbers and P is the property of primality, the binary number

$$N = .01110110110100110010\ldots$$

is equivalent to the list of odd primes

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, \ldots$$

A binary computer, with its ability to extract and examine a given binary digit, can use this compact method of storing information. The operation $N + N$ replaces N, which shifts the digits one place to the left and produces overflow if, and only if, the corresponding number is a prime, can be used in general to select the successive members of S having a property P. Other combinatorial processes involving several coded binary numbers can be used to advantage with a binary computer. For example, one can make a search for those objects of S that have a set of specified properties $P_1, P_2, \ldots$.

The binary system is implicit in a number of different arithmetical operations and games The so-called Russian peasant method of multiplying by doubling and halving is a case in point. To multiply 323 by 146, form two columns of figures (in the decimal system)

146	~~323~~
73	646
36	~~1292~~
18	~~2584~~
9	5168
4	~~10336~~
2	~~20672~~
1	41344
	47158

Each term of the first column is the integer part of half the preceding term. Each term of the second column opposite an even number in the first column is struck out. The sum of the remaining numbers gives the desired product 47158. The method works

classes *modulo* p.

$$\left(\frac{a}{p}\right)\cdot\left(\frac{b}{p}\right)=\left(\frac{ab}{p}\right)$$

If q is another odd prime then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}} \qquad (4)$$

the famous reciprocity law of quadratic residues. Moreover

$$\left(\frac{-1}{p}\right)=(-1)^{\frac{p-1}{2}} \qquad (5)$$

The Legendre symbol can be generalized to odd composite moduli (Jacobi symbol).

Reciprocity laws. For higher-power residues, these laws belong properly to algebraic number theory.

An expression

$$F(x,y) = Ax^2 + Bxy + Cy^2$$

is a binary quadratic form. Here the coefficients A, B, and C are taken as integers, the variables x and y assume only integer values. The expression, $D = B^2 - 4AC$, is called the determinant. Two forms

$$F(x,y) = Ax^2 + Bxy + Cy^2$$

and $\quad F_1(x',y') = A_1x'^2 + B_1x'y' + C_1y'^2$

are equivalent if there exists a linear transformation

$$x' = \alpha x + \beta y$$
$$y' = \gamma x + \delta y$$

with $\alpha\delta - \beta\gamma = \pm 1$, and α, β, γ, δ integers, so that

$$F(x,y) = F_1(x',y')$$

Equivalent forms have the same determinant and represent the same numbers. The number of classes of equivalent forms is finite. The theory is different for positive and negative determinants. There are only finitely many negative determinants which possess a given class number. In the theory of binary quadratic forms of positive determinant D, the so-called Pell's equation $t^2 - Du^2 = 1$ plays a fundamental role.

Pierre de F[illegible]
e [illegible]
[illegible]

[illegible] property is connected with the fact that for such a prime number -1 is a quadratic residue as Eq. (5) shows. J. L. Lagrange found that every natural number is the sum of at most 4 square[illegible]
E. W[illegible]
exi[illegible]
ber [illegible]
was [illegible] by David Hilbert by means of certain algebraic identities. Analytic proofs were later found by G. H. Hardy and J. E. Littlewood, and by I. M. Vinogradov.

There are infinitely many integral solutions for the Pythagorean equation

$$x^2 + y^2 = z^2$$

The solutions are all contained in the formula

$$x = u^2 - v^2 \qquad y = 2uv \qquad z = u^2 + v^2$$

where u and v are positive integers. Fermat made the famous statement that

$$x^n + y^n = z^n \qquad (6)$$

is not solvable in integers for any integral exponent $n > 2$. Besides $n = 4$, only prime exponents n need to be considered. Fermat's last theorem has been proved for all $n \leqq 4001$. In case x, y, z are prime to n, Eq. (6) is known to be unsolvable for all $n < 253,747,889$.

Algebraic number theory. The results concerning Fermat's last theorem are obtained through the methods of algebraic number theory. Karl F. Gauss carried over the concepts of number theory to the ring $R(i)$ of all complex integers $a + bi$, where a and b are ordinary integers. The law of unique prime factorization is preserved in this ring. Ordinary prime numbers $p \equiv 3 \pmod 4$ are also prime numbers in $R(i)$, whereas $2 = -i(1 + i)^2$ and the primes $p \equiv 1 \pmod 4$ are split $p = (a + bi)(a - bi)$. An algebraic number field $R(\theta)$ of degree n is generated by the root θ of an algebraic equation $F(x) = 0$ of degree n, having rational coefficients. A number α in this field is called an (algebraic) integer if it satisfies an algebraic equation with rational integer coefficients, the highest of which is 1. The algebraic integers in an algebraic number field form again an integral domain. However, the prime factorization is not necessarily unique, as the example $21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}) = (4 + \sqrt{-5})(4 - \sqrt{-5})$ shows, where each product is irreducible in the ring $R[\sqrt{-5}]$. Uniqueness is restored, after E. E. Kummer (1810–1893) and J. W. R. Dedekind (1831–1916) through the introduction of ideals. An ideal $\mathfrak{a}$ in the algebraic number field $K = R(\theta)$ is a set of algebraic integers of K, such that (i) if $\alpha \in \mathfrak{a}$, $\beta \in \mathfrak{a}$, then $(\alpha + \beta) \in \mathfrak{a}$, and (ii) if $\alpha \in \mathfrak{a}$ and ξ any algebraic integer in K, then $\alpha\xi \in \mathfrak{a}$. A multiplication of ideals can be defined, and then the concept of prime ideals can be introduced. A number α itself is in this theory replaced by the principal ideal (α), consisting of all numbers $\alpha\xi$, where ξ runs through all integers of K. The fundamental theorem is now the uniqueness of factorization of an ideal into prime ideals. Two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are called equivalent if two integers ξ and η in K exist such that $(\xi)\mathfrak{a} = (\eta)\mathfrak{b}$. Equivalent ideals are put into the same class. It turns out that the class number of K is finite. Kummer was led to the investigation of factorization in algebraic fields through his study of Fermat's Eq. (6), which can be written

$$(z - x)(z - \theta x)(z - \theta^2 x) \cdots (z - \theta^{n-1}x) = y^n \qquad (7)$$

where $\theta = e^{2\pi i/n}$ is a primitive nth root of unity. The factors of the left member of Eq. (7) are integers

Number theory

The primary objects of number theory are the properties and mutual relations of the natural numbers 1, 2, 3, . . . , and more generally of the integers, which include also the zero and the negative integers. The integers form a ring, that is, a domain within which addition, subtraction, multiplication (but not necessarily division) can always be carried out The numbers can be classified in many ways: for example, odd and even numbers, square numbers, prime numbers, perfect numbers. These classes of numbers were mentioned by Euclid (around 300 B C.).

In a wider sense, number theory also studies specific properties of other classes of numbers: rational, algebraic, and transcendental numbers. Several disciplines are distinguished in number theory.

Elementary number theory. The main concern of this branch of number theory is divisibility A number d is called a divisor of n (in symbols: $d \mid n$) if there exists an integer t such that $n = dt$. A prime number is a number that has only 1 and itself as divisors. Euclid proved that there is no last prime number in the sequence of integers Indeed, if 2, 3, 5, . . . , p are known prime numbers, then the number

$$N = 2 \cdot 3 \cdot 5 \cdot \cdots \cdot p + 1$$

is divisible by none of them and is therefore either a new prime number itself or is divisible by prime numbers not used in the construction of N. The fundamental theorem of number theory, also proved by Euclid, states that a number n can be factored into prime numbers in only one way when the order of the prime factors is disregarded The essential tool in the proof is Euclid's lemma: if a is a divisor of the product bc and if a and b are coprime, that is, they have no other common divisor than 1, then a divides c

A perfect number is defined as a number equal to the sum of its proper divisors, or divisors smaller than the number. Euclid showed that $(2^n - 1)2^{n-1}$ is a perfect number if $2^n - 1$ is a prime number, a so-called Mersenne prime. If $2^n - 1$ is a prime then n must itself be a prime. At present, 17 Mersenne primes and thus 17 perfect numbers are known, the first of which are 6, 28, 496. Euclid's formula yields only even perfect numbers. It is not known whether odd perfect numbers exist.

Congruences. If $a - b$ is divisible by m, then a is called congruent b modulo m, in symbols

$$a \equiv b \pmod m$$

This relation is an equivalence, that is, it is reflexive, symmetric, and transitive, and defines, therefore, equivalence classes of numbers congruent to each other, called residue classes There are evidently m residue classes modulo m. The number of those residue classes which contain only numbers coprime to m is called Euler's function $\varphi(m)$. Its value is

$$\varphi(m) = m \prod_{p \mid m} \left(1 - \frac{1}{p}\right)$$

Congruences to the same modulus can be subtracted, and multiplied in the same ma[illegible] equations. Two important congruences are

$$a^{p-1} \equiv 1 \pmod p$$

for p prime number, a not divisible by p (F[illegible] theorem), and Wilson's theorem

$$(p-1)! \equiv -1 \pmod p$$

Formula (1) was generalized by L. Euler to

$$a^{\varphi(m)} \equiv 1 \pmod m$$

where $(a,m) = 1$. The symbol (a,m) design[illegible] greatest common divisor of a and m.

The linear congruence

$$ax \equiv b \pmod m$$

is solvable for x if and only if $d \mid b$, whe[illegible] (a,m) Under this condition it has d differ[illegible] congruent) solutions modulo m. If $m =$ [illegible] prime number then

$$ax \equiv 1 \pmod p$$

has exactly one solution for each a not divi[illegible] p. This solution is the reciprocal of a mo[illegible] This shows that residue classes modulo p [illegible] form a ring, but also a field which is finite [illegible] p elements) and is of characteristic p (se[illegible] THEORY) The congruence

$$a_0x^n + a_1x^{n-1} + \cdots + a_n \equiv 0 \pmod{[illegible]}$$

is of degree n, if m does not divide a_0. If [illegible] then the number of solutions x cannot excee[illegible] a solution may not exist The congruence

$$x^{\varphi(m)} \equiv 1 \pmod m$$

has, in virtue of relation (2), $\varphi(m)$ solution[illegible] is a number coprime to m of which no lowe[illegible] than the $\varphi(m)$th is congruent 1 (mod m[illegible] called a primitive root modulo m. A mod[illegible] has primitive roots only for $m = 2, 4,$ [illegible] $m = p^k, 2p^k$, where p is an odd prime

The period of the periodic decimal into wh[illegible] reduced fraction l/m can be expanded [illegible] length $\varphi(m)$ or a divisor thereof It has th[illegible] mal length $\varphi(m)$ only if 10 is a primiti[illegible] modulo m

If the congruence

$$x^2 \equiv a \pmod m$$

is solvable, a is called a quadratic residue [illegible] m, otherwise a quadratic nonresidue. Let m [illegible] an odd prime. The Legendre symbol (a/p) [illegible] fined as $+1$ if a is a quadratic residue mo[illegible] and as -1 if a is a quadratic nonresidue, i[illegible] sumed that $a \not\equiv 0 \pmod p$. The Legendre [illegible] is a character of the multiplicative group of [illegible]

for example,

$$(5n+4) \equiv 0 \pmod 5 \quad p(7n+5) \equiv 0 \pmod 7$$

Diophantine approximations. It is evident that to any real number ω, one can find a rational number h/k with given denominator k which differs from ω by less than $1/k$. The theory of continued fractions shows that for certain denominators k, the approximation can be much better: for a given real irrational number ω, there exist infinitely many fractions h/k so that

$$\left|\omega - \frac{h}{k}\right| < \frac{1}{ck^2} \qquad (11)$$

Here any positive number $c \leqq \sqrt{5}$ can be chosen. For $c > \sqrt{5}$, however, there exist real numbers for which (11) has only finitely many solutions h/k. Also, the exponent 2 in (11) cannot be raised: for an irrational algebraic real ω the inequality

$$\left|\omega - \frac{h}{k}\right| < \frac{1}{k^{2+\epsilon}}$$

can have at most only finitely many solutions h/k (theorem of Thue-Siegel-Roth). Joseph Liouville *constructed* real numbers λ which for any given positive m have a solution h/k so that

$$\left|\lambda - \frac{h}{k}\right| < \frac{1}{k^m}$$

Such a number cannot be algebraic; it is transcendental. The numbers e and π were shown to be transcendental by Charles Hermite (1873) and Ferdinand Lindemann (1882), respectively. The latter result implies the impossibility of the quadrature of the circle.

If ω is irrational, then the fractional part of $n\omega$ approximates every real number r between 0 and 1, in other words: to a given $\epsilon > 0$, there exist infinitely many pairs of integers x and y such that $|x\omega - y - r| < \epsilon$. These fractional parts $\{r\omega\}$ are distributed uniformly in the interval $0 < r < 1$. That is to say, if α and β are two real numbers $0 \leqq \alpha < \beta < 1$ and $\Phi_N(\alpha,\beta)$, the number of $n \leqq N$ for which

$$\alpha \leqq \{n\omega\} \leqq \beta$$

then

$$\lim_{n\to\infty} \frac{1}{N} \Phi_N(\alpha,\beta) = \beta - \alpha$$

See ALGEBRA, NUMBER SYSTEMS, ZERO [H.R.]

Bibliography. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 2d ed., 1945; W. J. LeVeque, *Topics in Number Theory*, 2 vols., 1956; I. Niven, *Irrational Numbers*, Carus Mathematical Monograph 11, 1956.

Numerical analysis

The study of the methods and techniques for obtaining approximate solutions of apparently insoluble mathematical problems. Modern technology often requires the solution of mathematical problems which are so complex that only approximate solutions can be found.

Large systems of linear equations. The familiar methods of substitution and elimination for the solution of a system of n linear equations (that is, equations of the first degree) with n unknowns become quite unwieldly when the number of equations is large. Although in theory, a system of linear equations may always be solved with the aid of Kramer's rule (each unknown is the ratio of two determinants), in practice this method is just as unmanageable for large systems of equations as the methods mentioned above, so that numerical methods must be used. The standard procedure is to start with a first approximation to the solution of the given system of equations and to subject the approximate solution to a suitable process of refinement. The Gauss-Seidel method will illustrate the general procedure. Let the unknowns be designated by $x_1, x_2, x_3, \ldots, x_n$. To get a first approximation, all but the first term in the first member of the first equation is ignored, all but the first two terms in the first member of the second equation, and so on. In this manner, the resulting "triangular" system of equations can easily be solved and yields the approximate solutions $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \ldots, x_n^{(1)}$. To get a better approximation, one proceeds as follows: In the first equation, the unknowns $x_2, x_3, \ldots, x_n$ are replaced by the values $x_2^{(1)}, x_3^{(1)}, \ldots, x_n^{(1)}$, previously obtained, and the resulting equation is solved for x_1. Let this second approximation to x_1 be denoted by $x_1^{(2)}$. In the second equation, $x_3, x_4, \ldots, x_n$ are replaced by the values $x_3^{(1)}, x_4^{(1)}, \ldots, x_n^{(1)}$, and x_1 by $x_1^{(2)}$, and the resulting equation is solved for x_2. The value $x_2^{(2)}$ thus obtained is the second approximation to x_2, just as $x_1^{(2)}$ was the second approximation to x_1. Proceeding in this manner, one obtains in succession $x_3^{(2)}, x_4^{(2)}, \ldots, x_n^{(2)}$. The second set of approximate values is used in a manner similar to the above to obtain a third set of approximate values. It can be shown that the successive sets of approximate solutions converge to the true solutions of the given system of equations.

Numerical integration. The solution of numerous problems requires the evaluation of the area S under a given curve. If the given curve may be represented by the formula $y = f(x)$, the area under the curve and the x axis lying between the ordinates $x = a$ and $x = b$ is the definite integral of the function $f(x)$ between the limits $x = a$ and $x = b$, and is represented by the symbol

$$\int_a^b f(x)\,dx$$

If a function $F(x)$ may be found to have for its derivative the given function $f(x)$, then it is known that the desired area is merely the difference between the values of $F(x)$ at the upper and lower limits of the integral, namely b and a. Very often, however, the function $F(x)$, the so-called primitive of $f(x)$, does not have a simple formula, as, for in-

in the field $R(\theta)$. Ideals, first invented only for the use in number theory, have become a basic tool in higher algebra.

Analytic number theory. For certain problems of number theory, the methods of analysis, that is, of calculus and function theory, have to be used. The most famous problem of this sort is the number $\pi(x)$ of prime numbers $p \leq x$. About 1793, Gauss conjectured that

$$\lim_{x\to\infty} \frac{\pi(x) \log x}{x} = 1 \qquad (8)$$

This equation was proved a century later in 1896 by Jacques Hadamard and Charles J. de la Vallée Poussin. The methods were created by G. F. B Riemann (1826–1866) in a memoir of 1859 Riemann investigates the function $\zeta(s)$ defined by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad (9a)$$

Here $s = \sigma + it$ is a complex variable; the series in Eq. (9a) is convergent for $\sigma > 1$. By a method of analytic continuation, the function can be defined in the whole complex plane of s. It is a meromorphic function with only a simple pole of residue 1 at $s = 1$. On the other hand, the theorem of unique prime number factorization furnishes the representation

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}} \qquad (9b)$$

where p runs over all prime numbers. It can be shown that $\zeta(s)$ has no zeros for $\sigma = 1$. This fact and the pole at $s = 1$ suffice to establish Eq (8). Much finer statements about the function $\pi(x)$ have been proved Riemann's above-mentioned memoir contains the still unproved so-called Riemann hypothesis: all zeros of $\zeta(s)$ have a real part not exceeding ½ If this hypothesis is true, one would have

$$\pi(x) = \int_2^x \frac{dv}{\log v} + O(x^{(1/2)+\epsilon})$$

In 1949, the formula (8) was for the first time proved independent of the theory of the Riemann zeta function (ζ-function) by A. Selberg and P. Erdös.

Functions similar to $\zeta(s)$ were used by P G. L. Dirichlet (1837) to prove that there exist infinitely many prime numbers $p \equiv a \pmod m$, where a and m are given coprime numbers. Dirichlet defines a series

$$L(s\chi) = \sum_{n=1}^{\infty} \frac{\chi^{(n)}}{n^s} \qquad (10a)$$

where $\chi(n)$ is a character of the multiplicative group of residue classes modulo m and coprime to m. As a character modulo m, the function $\chi(m)$ satisfies the equations: $\chi(a_1) = \chi(a_2)$ for $a_1 \equiv a_2 \pmod m$ and $\chi(a)\chi(b) = \chi(ab)$. Here $\chi(n) = 0$ if n and m are not coprime. The series (10a) is convergent for $s > 1$. In virtue of the properties of the character χ, it permits also a multiplicative representation similar to Eq. (9b):

$$L(s,\chi) = \prod_p \frac{1}{1 - \chi^{(p)} p^{-s}} \qquad (10b)$$

where p runs over all prime numbers. There are $\varphi(m)$ characters χ, and thus so many functions $L(s,\chi)$. From the analytic behavior of $\log L(s,\chi)$ for the different χ if $s \rightarrow 1$, it can be deduced that

$$\sum_{p\equiv a \pmod m} \frac{1}{p^s} \rightarrow \infty \quad \text{if} \quad s \rightarrow 1$$

which implies that there must exist infinitely many primes $p \equiv a \pmod m$.

For problems of additive number theory, Euler utilized the simple fact $x^m \cdot x^n = x^{m+n}$ in his method of power series Let $m_1, m_2, m_3, \ldots$ be a monotone increasing sequence of nonnegative integers of a specified type (primes, squares, and so forth) The power series may be defined as

$$f(x) = x^{m_1} + x^{m_2} + x^{m_3} + \cdots$$

and form the qth power

$$f(x)^q = A_0 + A_1 x + A_2 x^2 + \cdots$$

Here the coefficients A_n signify the number of times that n can be formed as a sum of q elements from the set $\{m_i\}$. If all the A_n can be shown to be positive, then it is proved that every number n can be expressed as a sum of q numbers from the set $\{m_i\}$ For the determination of the coefficients A_n, functiontheoretical methods can be used, as was done by Hardy and Littlewood and later by Vinogradov. If the m_i are the kth powers and q is suitably large, a proof of Waring's theorem is obtained Taking the m_i as the successive odd prime numbers, this method led to the theorem that every sufficiently large odd number is the sum of at most 3 primes

A partition of n is the decomposition of n into additive parts, with disregard of order. For the number $p(n)$ of partitions of n, Euler gave the generating function

$$\sum_{n=0}^{\infty} p(n)x^n = \prod_{m=1}^{\infty} (1 - x^m)^{-1}$$

where $p(0) = 1$. This formula can be used to establish a recurrence formula for the computation of $p(n)$ Hardy and Srinivasa Ramanujan (1917) derived from it by functiontheoretical means an asymptotic expression for $p(n)$, the first term of which is

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi \sqrt{\frac{2n}{3}}\right)$$

Moreover, the number $p(n)$ can be expressed exactly as the sum of a convergent infinite series The arithmetical function $p(n)$ possesses various congruence properties, discovered by Ramanujan.

order differences gives rise in succession to a column of second-order differences containing $n-1$ entries, a column of third order differences containing $n-2$ entries, and so on. For any one argument $x_k = x_0 + kh$, the differences of various order slanting diagonally downward are referred to as the differences corresponding to the entry $y_k = f(x_k)$. These differences are denoted by the symbols Δy_k, $\Delta^2 y_k$, $\Delta^3 y_k$, ... The most important formula involving differences of ascending order is the so-called Newton-Gregory interpolation formula

Let it be desired to calculate the value of $f(x)$ for $x = x_0 + ph$, where p is smaller than unity. Then the Newton Gregory formula reads

$$f(x_0+ph) = y_0 + \frac{p}{1}\Delta y_0 + \frac{p(p-1)}{1\cdot 2}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{1\cdot 2\cdot 3}\Delta^3 y_0 + \cdots + \Delta^n y_0$$

(The coefficients in the above formula are the coefficients of the binomial expansion formula.) Since the values of the differences either are listed in the table of the given function or, at any rate, are readily obtainable by a series of simple operations of subtraction, it is seen that the evaluation of $f(x_0+ph)$ for various values of p requires the evaluation of the coefficients

$$\frac{p}{1}, \frac{p(p-1)}{1\cdot 2}, \frac{p(p-1)(p-2)}{1\cdot 2\cdot 3}, \quad \ldots$$

for the same values of p. The amount of labor involved in the computation of these coefficients is considerably less than that involved in the evaluation of the Lagrangian interpolation polynomial. An additional advantage of the Newton-Gregory interpolation formula arises from the fact that quite frequently the terms involving differences of high order are small and may therefore be neglected.

Many problems in advanced physics and engineering are considerably more complex than those discussed above. Most of these problems require the solution of differential equations, that is, equations involving derivatives of unknown functions. The standard numerical method for solving such differential equations is to replace the various derivatives by the appropriate difference quotient. Thus, as is known to any student of calculus, the first derivative of a function $f(x)$ is defined as the limit of the quotient

$$\frac{f(x+h) - f(x)}{h}$$

as h approaches zero. Similarly, it is shown in calculus that the second derivative of a function is the limit of the quotient

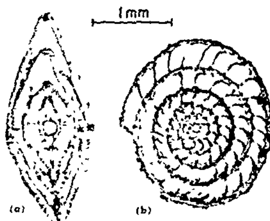
$$\frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

as h approaches zero. Thus, when solving a differential equation involving the first and second derivatives numerically, these derivatives are replaced by the quotients just mentioned. The resulting equation makes it possible to evaluate the solution of the given differential equation when certain boundary conditions are given. It may be mentioned briefly that finite differences, as defined here, play a major role in the solution of differential equations. See COMPUTER; DETERMINANT; GRAPHIC METHODS; INTEGRATION; INTERPOLATION; POLYNOMIAL SYSTEMS OF EQUATIONS. [A.N.L.]

Bibliography: K. S. Kunz, *Numerical Analysis*, 1957.

Nummulites

An extinct genus of the relatively large (up to 35 mm in diameter), unicellular protozoans of the order Foraminifera (properly known as *Camerina*). Individuals of this genus developed a lenticular, discoidal, planispirally coiled, involute, multichambered test (shell), composed of calcium carbonate. They inhabited shallow, warm, marine waters of the tropical zone during the early Tertiary period. See FORAMINIFERA FOSSILS.



Photomicrographs of the internal structure of Nummulites. (a) Transverse section (b) Median section. (After W. S. Cole)

Although *Nummulites* occurs in the Eocene rocks of the Americas, its maximum development was during the Eocene and Oligocene epochs in the circum-Mediterranean region. There, individuals were so abundant that certain limestones are composed almost entirely of their tests. The pyramids of Egypt are made of blocks of an Eocene nummulitic limestone. The numerous species are used in geologic correlation. See LIMESTONE [W.S.C.]

Nursing

The application of the principles of physical, biological, and social sciences in the physical and mental care of people, sick and well. Nursing includes therapy as directed by the doctor, physical and emotional care; and patient and family education in rehabilitation, health maintenance, and disease prevention. When several persons are simultaneously involved in the nursing care of a person or

stance, if $f(x)$ is e^{-x^2}. In such case numerical methods must be used to evaluate the given definite integral. A brief discussion of several methods for obtaining an approximation to the value of the desired integral will illustrate the general approach of numerical analysis in dealing with "insoluble" mathematical problems. Assume, for the sake of completeness, that the value of $f(x)$ has been evaluated for a total of $n + 1$ equidistant values of x, including $x = a$ and $x = b$. If these values of x are $x_0 = a$, $x_1 = a + h$, $x_2 = a + 2h$, . . . , $x_n = a + nh = b$, and if the corresponding points on the curve $y = f(x)$ are denoted by $A_0, A_1, A_2, \ldots, A_n$, and their ordinates by $y_0, y_1, y_2, \ldots, y_n$, then the simplest estimate of the desired integral is obtained by replacing the curve $y = f(x)$ by the polygonal line passing through the points $A_0, A_1, A_2, \ldots, A_n$ and calculating the area under the polygonal line by breaking it up into n trapezoidal strips. Since a typical trapezoidal strip has the area $\frac{1}{2}h(y_k + y_{k+1})$, one readily obtains the approximate integration formula (or mechanical integration formula)

$$S \cong h\left(\frac{a}{2} + y_1 + y_2 + \quad + y_{n-1} + \frac{b}{2}\right)$$

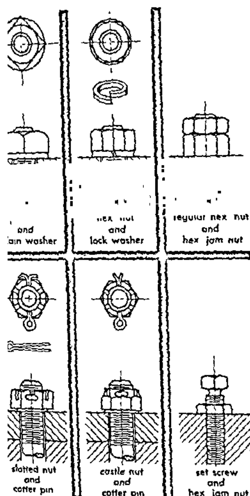
It is conceivable that a more accurate approximation to the desired area may be obtained by replacing the portion of the curve corresponding to two consecutive intervals of combined length $2h$ by a portion of a parabola. Since the expression for the area under a parabola is known from calculus, the desired approximation is obtained by adding the expressions for the areas of the strips bounded by arcs of parabolas, included in the integral (a, b). Still another approximation to the desired area may be obtained by replacing the portion of the curve $y = f(x)$ corresponding to three consecutive intervals of combined length $3h$ by a portion of a cubic Since the expression for the area under a cubic is also known, the desired new approximation is obtained by adding the expressions for the strips bounded by arcs of cubics included in the given interval (a, b). In this discussion, it is assumed that $f(x)$ is known, but its primitive $F(x)$ is not It is obvious that the technique just described is equally applicable to the case where the mathematical expression for $f(x)$ is actually not known, even though the quantities above designated by $y_0, y_1, y_2, \ldots, y_n$ are known (very often these quantities are the results of physical measurements). Since the formula for the area under the curve representing a given polynomial is known from the calculus, it is reasonable to conjecture that when $n + 1$ equidistant ordinates of a given curve are known, the optimum solution of the problem of evaluating the definite integral under consideration is to obtain the so-called Lagrangian polynomial which assumes the known values $y_0, y_1, y_2, \ldots, y_n$ at the given $n + 1$ equidistant values of x and to use the known calculus formula for the primitive of a given polynomial.

Interpolation. Suppose, again, the values of function $f(x)$ are known at $n + 1$ equidistant [illegible] ments (that is, values of x) and that it is desir[illegible] calculate the value of $f(x)$ for some interme[illegible] argument. For the sake of concreteness, let [illegible] desired to calculate $f(x)$ for a value of x [illegible] between $x_0 = a$ and $x_1 = a + h$. The simplest [illegible] least accurate, procedure is to interpolate lin[illegible] between the known values y_0 and y_1. This proc[illegible] is, of course, tantamount to replacing the arc [illegible] necting the points of ordinates y_0 and y_1 [illegible] straight line. In the light of the previous discu[illegible] of numerical integration, it is clear that a mor[illegible] curate interpolation procedure would be to re[illegible] the portion of the curve $y = f(x)$ correspond[illegible] the three points whose ordinates are y_0, y_1, a[illegible] by an arc of a parabola, and to calculate th[illegible] sired value by means of the formula, known [illegible] the calculus, for the parabola passing thr[illegible] three given points corresponding to equidista[illegible] guments. Further and more accurate proce[illegible] would be to use the formula for the cubic pa[illegible] through four points corresponding to equid[illegible] arguments or, better yet, the Lagrangian for[illegible] for the polynomial $P_n(x)$ passing through [illegible] points corresponding to $n + 1$ equidistant [illegible] ments. It can be readily verified that

$$P_n(x) = \frac{(x - x_1)(x - x_2) \cdot \quad \cdot (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \cdot \cdot \quad (x_0 - x_n)} y_0 + \frac{(x - x_0)(x - x_2) \quad (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \quad (x_1 - x_n)} \cdot y_1 + \quad + \frac{(x - x_0)(x - x_1) \quad \cdot (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \quad (x_n - x_{n-1})}$$

It is seen from this formula that the evaluatio[illegible] $P_n(x)$ for a given value of x requires the evalu[illegible] of $n + 1$ expressions (the numerators of the n [illegible] fractions), each one of which is a product of n [illegible] tors It is clear that whenever the value of P_n([illegible] desired for some new value of x, the $n + 1$ proc[illegible] must be recomputed from the beginning. T[illegible] while the Lagrangian interpolation formul[illegible] characterized by high accuracy, it has the [illegible] disadvantage of being uneconomical whenev[illegible] must be used for a large number of arguments [illegible] this reason, other interpolation schemes have [illegible] designed which obviate the above disadvantage [illegible] very marked degree.

To describe the structure of such interpola[illegible] schemes, it is necessary to define finite differe[illegible] of various orders Assume that values of a func[illegible] $f(x)$ corresponding to a sequence of equidis[illegible] arguments are listed in a vertical column along [illegible] the column of the arguments. If, from each tab[illegible] value, the value corresponding to the prece[illegible] argument is subtracted, the values thus obtai[illegible] are called the differences of the first order. Fro[illegible] column containing $n + 1$ entries, one evide[illegible] obtains a column of first-order differences cont[illegible] ing n entries (each entry being written midway [illegible] tween the two tabular values from which it [illegible] obtained). In a similar manner, the colum[illegible]



pplications of nuts.

orms such as jam nut, castellated nut, slotted nut, ap nut, wing nut, and knurled nut (see illustration)

Hexagon jam nuts are used as locking devices to keep regular nuts from loosening and for holding et screws in position. They are not as thick as plain nuts. Jam nuts are available in semifinished form in both regular and heavy weight.

Castellated and slotted nuts have slots so that a cotter pin or safety wire can hold them in place. They are commonly used in the automotive and allied fields on operating machinery where nuts tend to loosen. The slotted nut is a regular hexagon nut with slots cut across the flats of the hexagon. They are standardized in regular and heavy weights in semifinished hexagon form. Finished thick slotted nuts are available. Castle nuts are hexagonal with a cylindrical portion above through which slots are cut.

Machine screw and stove-bolt nuts may be either square or hexagonal. Hexagon machine-screw nuts may have *the top chamfered* at 30° with a plain *bearing surface*, or both top and bearing surfaces may be chamfered. Square nuts have flat surfaces without chamfer. Square nuts are available with a coarse thread; hexagon nuts may be supplied with either coarse or fine thread.

Wing and knurled nuts are designed for applications where a nut is to be tightened or loosened using finger pressure only. *See* SCREW FASTENER. [W.J.L.]

Nut crop culture

The cultivation of *nut* crops, which popularly include not only the *true* nuts, such as walnuts and pecans, but also many other dry fruits or seeds with hard shells and interior kernels or meat, such as coconuts [illegible] formed from basal appendages of the flower or from the involucre.

Nuts are concentrated sources of energy with high food value for man and animals. Many kinds contain 50–70% fats and oils and 15–30% protein in the dried kernels.

Nuts gathered from wild trees still make up a large part of the world production, especially of

Nuts important in world commerce

Common name	Kind of plant	Origin	Climatic zone adaptation; principal current sources	Estimated total annual commercial production of principal producing countries of record, short tons	Principal uses and producing areas
Acorn	Deciduous tree	Europe, United States, Asia	Temperate, wild	Data unavailable	Animal feed, Europe, United States, Asia
Almond[a]	Deciduous tree	Asia Minor	Temperate, subtropical, planted	100,000 (shelled)[a]	Food; Mediterranean countries, United States, China, Iran
Brazil nut[a]	Evergreen tree	South America	Tropical; wild	30,000 (unshelled)[a]	Food; Brazil, Bolivia

persons, nursing may also include the application of the principles and skills involved in management

History. Modern nursing began in 1860 when Florence Nightingale established the school which bears her name at St. Thomas's Hospital, London, England. Linda Richards, called the first trained nurse in America, was graduated from the New England Hospital for Women and Children in Boston, Massachusetts, in 1872. Three training schools for nurses, opened on the Nightingale plan in 1873, laid the foundation for nursing education in the United States. These schools were the Bellevue at the Bellevue Hospital in New York, the Connecticut at the New Haven Hospital in New Haven, and the Boston, now the Massachusetts General Hospital School in Boston.

Legally two types of nurses are recognized, the registered nurse, commonly referred to as the professional nurse, and the licensed practical nurse. After graduation from a state-approved school of nursing, a written examination set by the state authority is required before legal recognition can be obtained either by nurse or practical nurse.

Nursing in the United States has remained as it began, largely an occupation for women. Approximately 2.4% of the practicing registered nurses are men, who numerically fill a small but important role in all fields of nursing. The improved status of women, the growing educational opportunities for women, the demands for greater technical, social, and judgmental skill and leadership in the practice of nursing have been evident in the changing pattern of education. For many years, based largely on the apprenticeship system of learning, modern professional nursing education has moved gradually toward the educational patterns of other professions

Nursing has long been organized. The American Nurses' Association, founded in 1896, the professional organization for nurses, furnishes a medium for action relating to the employment of nurses and the improvement of nursing practice. This organization was a charter member of the International Council of Nurses, founded in 1899, through which the American nurses can contribute to the improvement of the care of the sick throughout the world. In 1952, the National League of Nursing Education, 1893, the National Organization for Public Health Nurses, 1912, and the Association of Collegiate Schools of Nursing, 1935, merged into a National League for Nursing, a significant social experiment in organization. The National League for Nursing brings together nurses, practical nurses, and others concerned in nursing care, members of allied professions, and citizen consumers of nursing services to work for the development and improvement of nursing services in hospitals, public-health agencies, schools, and industries, and of all types of education for nursing.

Types of school. Three types of school prepare nurses for beginning practice: the 3-year or diploma program in the school conducted by a hospital; the 2-year or associate in arts degree program, conducted by the junior or community college; and the 4- and 5-year or baccalaureate degree programs, conducted by the senior college or university. All three types of programs admit applicants directly after high school graduation, and prepare for direct patient care in general, mental, obstetrical, children's, and communicable disease hospitals, clinics, and operating rooms The baccalaureate program prepares also for beginning practice in public health or visiting nursing Specialization in any field of nursing, and preparation for teaching, supervision, and administration, begins in the master's degree programs offered by the university. Preparation for research, or teaching in the university schools of nursing, is secured throug further study in the university for the doctora degree.

The first school for practical nurses in this coun try was established in 1912 in New York City. Al though the early schools demonstrated the value o the practical nurse, this plan of training develope slowly until World War II. Few schools have bee established as independent schools since the b ginning of this training Schools for practic nurses are now conducted by hospitals and by th state education systems in the vocational hig schools, which secure facilities for bedside trainin in cooperating hospitals. The program for th licensed practical nurse is commonly 1 year length, and prepares for the care of the mildly i in the home, hospital, and nursing home, under i direct supervision of doctor or nurse, or for th care of the more acutely ill patients under dire supervision

Role in government health services. Professional nurses play a major role in the government health services They serve in the military nurse corps of the Army, Navy, Air Force, and Public Health Service as commissioned officers As civilian nurses under Civil Service or other government programs, they are employed by the Bureau of Indian Affairs, the Public Health Service, and the Veterans Administration Department of Medicine and Surgery Nurses hold positions with large administrative and consultative responsibilities in such agencies as the Children's Bureau, the Civil Service Commission, Office of Civilian Defense Mobilization, and in foreign service under the United States International Cooperation Administration. In natural disaster or national emergency, the nurse may serve as a member of the American Red Cross Nursing Service. *See* MEDICINE. [R.SL.]

Nut

In mechanical structures, an internally threaded fastener. Plain, square, and hexagonal nuts for bolts and screws are available in three degrees of finish: unfinished, semifinished, and finished There are two standard weights: regular and heavy. For specific applications, there are other standard

...e angular frequency of the nutation of a top ... at a high spin rate is given by

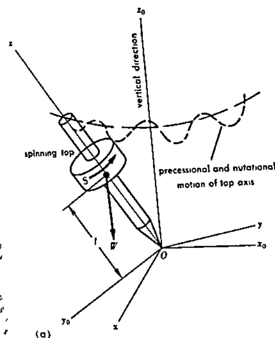
$$\omega_n = \frac{I_z}{I_x} S \qquad (1)$$

...ere I_z and I_x are moments of inertia about the ...nd x axes, respectively, and S is the angular ve-...ity of spin Furthermore, the rate of precession ... for the general motion is not uniform but varies ...rmonically with time with the same frequency as ...es the nutation·

$$\omega_p = \frac{Wl}{I_z S}(1 - \cos \omega_n t) \qquad (2)$$

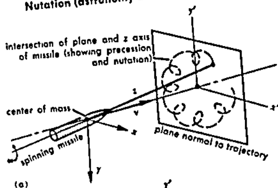
The average precessional frequency is then

$$(\omega_p)_{ave} = \frac{Wl}{I_z S} \qquad (3)$$



(b) (c) (d)

Fig 1 Motion of spinning top, showing typical traces of the top spin axis on unit spheres for different initial conditions (a) General motion (b) Top released with initial angular velocity in direction of precession (c) Top released with initial angular velocity opposite to direction of precession. (d) Axis of spin released from rest



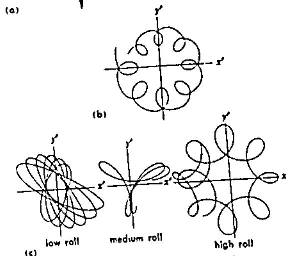


Fig 2. Precessional and nutational displacements of missiles in flight (a) Sketch showing intersection of missile axis on a plane The velocity of the center of mass of the missiles is indicated by v (b) Typical motion of spin stabilized missiles (bullets, shells, and the like) (c) Typical motion of rolling fin-stabilized missiles for low, medium, and large rates of roll, respectively

As the spin rate S is increased the frequency of nutation increases, as shown by Eq. (1), and the nutational displacement $(\theta_2 - \theta_1)$ decreases very rapidly. Furthermore, as S is increased in Eq. (2), the frequency of the precessional variation increases, but from Eq (3) the average rate of pre-... [illegible] ...

uniformly about the vertical axis for this common case

Nutation of projectiles. The motion relative to the centers of mass of bullets and shells stabilized by high rates of spin is identical to that of a spinning top relative to the fixed point of contact Torques about the centers of mass due to aerodynamic forces acting on such bodies during flight cause precession and nutation to occur (Fig 2). Furthermore, finned missiles, which usually rotate or spin rather slowly during flight unless prohibited from doing so by a suitable control system, also develop precessional and nutational angular velocities, similar to a spinning pendulum [R E.BO.]

Astronomical nutation. The rotating Earth can be regarded as a spinning symmetrical top with

Nuts important in world commerce (Cont.)

Common name	Kind of plant	Origin	Climatic zone adaptation; principal current sources	Estimated total annual commercial production of principal producing countries of record, short tons	Principal uses and producing areas
Cashew[a]	Evergreen tree	Tropical America	Tropical; planted	160,000 (unshelled)[b]	Food, commercial oil; India, East Africa
Chestnut[a]	Deciduous tree	Asia Minor, China, Japan	Temperate, planted	Data unavailable	Food; Mediterranean countries, China, Japan
Coconut[a]	Palm tree	Probably Polynesia	Tropical; planted and wild	3,300,000 (copra)[c]	Food, edible and commercial oil, soap, fiber, thatching; Philippines, Ceylon, Malaya, Indonesia
Cola nuts[a]	Evergreen tree	Africa	Tropical; planted and wild	Data unavailable	Beverages, masticatory, food, West Africa
English or Persian walnut[a]	Deciduous tree	Asia Minor	Temperate, subtropical; planted	130,000 (unshelled)[b]	Food, oil; Mediterranean countries, United States, China, India
Filbert	Deciduous shrub or small tree	Asia Minor	Temperate, planted	130,000 (unshelled)[b]	Food; Turkey, Italy Spain, United States, Afghanistan
Palm kernels[a] and palm oil	Palm tree	West Africa	Tropical, wild and planted	1,000,000 (kernels)[c] 1,100,000 (oil)[c]	Edible and commercial oil, food, West Africa, Indonesia, Brazil
Peanut[a]	Annual plant	South America	Tropical, subtropical, warm temperate; planted	14,300,000 (unshelled)[c,d]	Edible oil, food; India, Africa, China, United States
Pecan	Deciduous tree	North America	Temperate, planted and wild	70,000 (unshelled)[b]	Food, United States
Pine nuts	Evergreen tree	Europe, Asia, North America	Temperate, wild	Data unavailable	Food, Europe, Asia North America
Pistachio[a]	Deciduous tree	Asia Minor	Temperate, planted and wild	Data unavailable	Food, Iran, Turkey, Syria, Italy
Tung[a]	Deciduous tree	China	Warm temperate, planted	150,000 (oil)[b]	Oil, paints, China, United States

[a] Separate articles given under common name [b] USDA *Agr Statistics*, 1958 [c] FAO Yearbook, 1957
[d] Excluding U.S.S.R.

tropical kinds. However, most nuts are now produced in well-cared-for plantings where improved varieties and modern cultural methods are used. Important nuts in world commerce are described in the table; separate articles given under common names are indicated by footnote. *See* COLA; OAK; PALM; PINE. [E.F.S.]

Nutation (astronomy and mechanics)

In mechanics, the term nutation refers to a bobbing or nodding up-and-down motion of a spinning rigid body, such as a top, as it precesses about its vertical axis. Astronomical nutation refers to irregularities in the precessional motion of the equinoxes caused by the varying torque applied to the Earth by the Sun and Moon. Astronomical nutation, sometimes called nutational wandering of the terrestrial poles, should not be confused with nutation as defined in mechanics; the latter is present even if the source of the torques is unvarying.

Nutation of tops. The general motion of a spinning top, easily observed at low spin rates, consists of both precession and nutation (*see* PRECESSION). Figure 1*a* shows a symmetrical top spinning about a fixed point with its axis tracing out this general motion. Figure 1*b* shows the motion for the case when the axis of the spinning top is released with an initial angular velocity in the direction of precession; Fig. 1*c*, that with an initial velocity opposite to the precession; and Fig. 1*d*, that with zero initial angular velocity (axis of spin released from rest).

evergreen 30–60 ft high, a member of the nutmeg family (Myristicaceae) The golden-yellow, ma-re fruits resemble apricots They gradually lose oisture and when completely ripe, the husk (peri-irp) splits open exposing the shiny brown seed overed with a red, fibrous, aromatic aril which is ie mace The kernel inside the seed coat is the nut-ieg of commerce. Fruits are produced throughout he year and are picked when the husks split open The mace is removed from the husks, flattened, and dried It is used in making pickles, ketchup, and sauces When the seeds are thoroughly dried the shells are cracked off, the kernels are removed, sorted, and often treated with lime to prevent dam-age by insects Grated nutmeg is used in custards, puddings, and other sweet dishes; also in various verages Nutmeg oil is used in medicine, per-mery, dentifrices, and in the tobacco industry e RANALES; SPICE AND FLAVORING. [F.O.S.]

utrition

he science of food, concerned with the processes

ouy tissues, the liberation of energy required to maintain body temperature, to sustain respiration, heartbeat, and other functions essential to life; and he provision of energy for exercise and work

Nutrients are generally grouped into a few large categories, depending on their chemical formulas or characteristics and their functions in the body. The six major categories are proteins, carbohydrates, lipids, vitamins, minerals, and water.

Proteins always contain nitrogen, as well as carbon, hydrogen, oxygen, and generally sulfur Many contain phosphorus, and elements such as iodine, iron, copper, and zinc are also occasionally present Proteins are highly complex substances consisting of amino acids linked together. Of the 20–25 known amino acids, 8 (L-tryptophan, L-phenylalanine, L-lysine, L-threonine, L-valine, L methionine, L-leucine, L-isoleucine) are essential for man They can not be synthesized by the body, so they must be provided by the diet. The proteins are the principal nitrogenous constituents of all tissue and play an important role in body processes See AMINO ACIDS, PROTEIN

The carbohydrates contain carbon, hydrogen, and oxygen The major function of carbohydrates is to provide energy, they contribute from half to two-thirds of man's energy under normal conditions. They also have a sparing effect on protein, that is, proper utilization of carbohydrate means that protein is not being used for this purpose, therefore tissue is not broken down In addition, carbohydrates improve the utilization of fat. See CARBOHYDRATE.

Lipids consist of the compounds known as fats, waxes, phospholipids, glycolipids, and sterols, as well as some of their hydrolytic products They are composed chiefly of carbon, hydrogen, and oxygen, although other elements may be present Probably the most important role of lipids is that of fuel; they yield more than twice as much energy per gram as carbohydrate or protein. Lipids also delay the onset of hunger, since they prolong the time food stays in the stomach They provide the essential fatty acids, linoleic, linolenic, and arachidonic and are vehicles by which the fat-soluble vitamins (A, D, E, and K) enter the body; lipids also spare the need for thiamine See LIPID; VITAMIN.

Vitamins, water, and certain inorganic substances must also be included in the diet. Many of the trace elements function as essential components of enzyme systems. Deficiencies or beneficial effects of some elements, such as calcium, phosphorus, iron, iodine, fluorine, sodium, and potassium have been observed in man. Other elements, including copper, molybdenum, zinc, and cobalt, are presumed to be important to man, although deficiencies have not been demonstrated. See BIOCHEMISTRY. [F.J.ST.]

Nylon

A polyamide resin. In textiles, monofilament nylon is used to make hosiery and filter cloths; multifilament yarns and nylon staple, either alone or blended with other yarns, are used to make a wide variety of fabrics Thicker nylon extrusions are used to make tennis racket strings, rope and line, surgical sutures, and paintbrush bristles; and the nylon raw material can also be made into injection plastic parts, such as zipper teeth and tubing See POLYAMIDE RESIN. [C CO.]

Nystatin

An antifungal antibiotic useful in the therapy of a wide variety of nonsystemic fungal infections and also as an ingredient in animal feeds for enhanced growth rates with poultry and swine. It is active against a wide range of yeasts and other fungi including *Candida*, *Aspergillus*, *Penicillium*, *Botrytis*, and others, but it is without activity against bacteria. Chemically it is a polyene (see AMPHOTERICIN B). It is produced biosynthetically by fermentation with a strain of *Streptomyces noursei* In 1956 more than 22,000 lb was produced in the United States by the only manufacturer. See DERMATOPHYTOSIS

Chemistry. The chemical structure of nystatin has not been determined. It has the characteristic ultraviolet absorption spectrum of a conjugated tetraene. It is ...
with the t...
with a m...
an activit... units/mg It is insoluble in ether, chloroform, and acetone, very slightly soluble in water, methanol, ethanol, butanol, or dioxane; soluble in dimet...
amide,
Finely ...
laborat... ...splay may be prepared by slowly pouring concentrated solutions into large quantities of water.

Assay and microbial activity. Nystatin may be assayed for biological activity by a diffusion plate

small angular speed but large angular momentum, the latter due to its large mass. As detailed in another article (*see* PRECESSION OF EQUINOXES), the gravitational attractions of the Sun and Moon cause the Earth's axis to describe a cone about the normal to the plane of its orbit. However, the magnitude of these gravitational attractions is continually varying, due to the changing positions in space of Sun, Moon, and Earth. The Moon's orbit is continually changing its position in such a way that the celestial pole undergoes a nodding (nutation) as well as a periodic variation in the rate of advance. The largest nutation is about 9″.2, and occurs in a period of a little less than 19 years; that is, the celestial pole completes a small ellipse of semimajor axis 9″.2 in about 19 years.

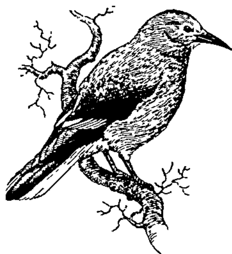
There are lesser nutation effects which are due to the motion of the Moon's nodes, the changing declination of the Sun, and the changing declination of the Moon.

Nutation of gyroscopes. Still another example of the nutation of a spinning symmetrical body is given by the general motion of a gyroscope. For a discussion of this, *see* GYROSCOPE. [K.W.P.]

Bibliography: L. Davis, Jr., J. W. Follin, and L. Blitzer, *Exterior Ballistics of Rockets*, 1958; H. Goldstein, *Classical Mechanics*, 1950

Nutcracker

Either of two species of the genus *Nucifraga*. One species occurs in Eurasia, the other in western North America. They belong to the family Corvidae, and in general features are much like the other crows. Clark's nutcracker, *N. columbiana*, ranges in the mountains from Alaska to Mexico. It has a



Clark's nutcracker, *Nucifraga columbiana*, length to 13 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

light gray body, black wings and tail, and conspicuous white patches on both the wings and tail. Nutcrackers are noisy, inquisitive visitors of camps, and become quite tame when encouraged and protected, as in some national parks of the United States. *See* PASSERIFORMES. [J.D.B.]

Nuthatch

Any member of the family Sittidae, a family related to the chickadees. There are 29 known species, of which 4 occur in the United States. The nuthatches are small, chunky creepers; they scan the surface of tree trunks and limbs, and are the only birds which regularly descend head downward. All American species have gray backs. Largest and most common is the white-breasted nut-



The white-breasted nuthatch, *Sitta carolinensis*, length 6⅛ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

hatch, *Sitta carolinensis*, which occurs from southern Canada throughout the United States southward into central Mexico. The male has a black crown and nape with a white stripe above the eye; the crown is gray on the female. *See* PASSERIFORMES. [J.D.B.]

Nutmeg

A delicately flavored spice obtained from the nutmeg tree, *Myristica fragrans*, a native of the Moluccas, or Spice Islands. The tree is a dark-leaved



Nutmeg (*Myristica fragrans*) (USDA)

ergreen 30–60 ft high, a member of the nutmeg mily (Myristicaceae). The golden-yellow, mare fruits resemble apricots They gradually lose oisture and when completely ripe, the husk (pericarp) splits open exposing the shiny brown seed overed with a red, fibrous, aromatic aril which is he mace The kernel inside the seed coat is the nutmeg of commerce Fruits are produced throughout he year and are picked when the husks split open The mace is removed from the husks, flattened, and dried It is used in making pickles, ketchup, and sauces. When the seeds are thoroughly dried the shells are cracked off, the kernels are removed, sorted, and often treated with lime to prevent damage by insects. Grated nutmeg is used in custards, puddings, and other sweet dishes, also in various beverages Nutmeg oil is used in medicine, perfumery, dentifrices, and in the tobacco industry. *See* RANALES; SPICE AND FLAVORING. [P D S]

Nutrition

The science of food . . .

body tissues, the liberation of energy required to maintain body temperature, to sustain respiration, heartbeat, and other functions essential to life; and the provision of energy for exercise and work.

Nutrients are generally grouped into a few large categories, depending on their chemical formulas characteristics and their functions in the body. The six major categories are proteins, carbohydrates, lipids . . .

. . . phosphorus, and elements such as iodine, iron, copper, and zinc are also occasionally present Proteins are highly complex substances consisting of amino acids linked together Of the 20–25 known amino acids, 8 (L-tryptophan, L-phenylalanine, L-lysine, L-threonine, L-valine, L-methionine, L-leucine, L isoleucine) are essential for man. They cannot be synthesized by the body, so they must be provided by the diet The proteins are the principal nitrogenous constituents of all tissue and play an important role in body processes *See* AMINO ACIDS; PROTEIN

The carbohydrates contain carbon, hydrogen, and oxygen. The major function of carbohydrates is to provide energy; they contribute from half to two-thirds of man's energy under normal conditions They also have a sparing effect on protein; that is, proper utilization of carbohydrate means that protein is not being used for this purpose; therefore tissue is not broken down In addition, carbohydrates improve the utilization of fat *See* CARBOHYDRATE.

Lipids . . .

the most important role of lipids is that of fuel; they yield more than twice as much energy per gram as carbohydrate or protein. Lipids also delay the onset of hunger, since they prolong the time food stays in the stomach. They provide the essential fatty acids, linoleic, linolenic, and arachidonic and are vehicles by which the fat soluble vitamins (A, D, E, and K) enter the body; lipids also spare the need for thiamine. *See* LIPID; VITAMIN.

Vitamins, water, and certain inorganic substances must also be included in the diet. Many of the trace elements function as essential components of enzyme systems. Deficiencies or beneficial effects of some elements, such as calcium, phosphorus, iron, iodine, fluorine, sodium, and potassium have been observed in man. Other elements, including copper, molybdenum, zinc, and cobalt, are presumed to be important to man, although deficiencies have not been demonstrated. *See* BIOCHEMISTRY. [F.J.ST.]

Nylon

A polyamide resin. In textiles . . .

with other yarns, are used to make a wide variety of fabrics. Thicker nylon extrusions are used to make tennis racket strings, rope and line, surgical sutures, and paintbrush bristles; and the nylon raw material can also be made into injection plastic parts, such as zipper teeth and tubing. *See* POLYAMIDE RESIN. [C CO]

Nystatin

An antifungal antibiotic useful in the therapy . . . wide variety . . . also as an ing . . . growth rates with poultry and swine It is active against a wide range of yeasts and other fungi including *Candida*, *Aspergillus*, *Penicillium*, *Botrytis*, and others, but it is without activity against bacteria Chemically it is a polyene (*see* AMPHOTERICIN B). It is produced biosynthetically by fermentation with a strain of *Streptomyces noursei*. In 1956 more than 22,000 lb was produced in the United States by the only manufacturer. *See* DERMATOPHYTOSIS

Chemistry. The chemical structure of nystatin has not been determined It has the characteristic ultraviolet absorption spectrum of a conjugated tetraene. It is a crystalline, light-yellow substance with the tentative empirical formula of $C_{46}H_{75}NO_{18}$ with a molecular weight of 930 Pure nystatin has an activity of 6000 units/mg. It is insoluble in ether, chloroform, and acetone; very slightly soluble in water, methanol, ethanol, butanol, or dioxane, soluble in dimethylsulfoxide, dimethylformamide, and in 1–2% $CaCl_2$ in anhydrous methanol. Finely dispersed suspensions in water for use in the laboratory or as a plant spray may be prepared by slowly pouring concentrated solutions into large quantities of water.

Assay and microbial activity. Nystatin may be assayed for biological activity by a diffusion plate

assay using *Candida albicans* or *Saccharomyces cerevisiae* as a test organism. *See* BIOASSAY.

The antimicrobial activity of nystatin is fungicidal and is confined to the yeasts and fungi; it has *no activity against the bacteria and actinomycetes* including *Streptomyces*.

In vitro inhibition at concentrations between 1 and 3 units/ml is obtained with *Candida albicans, Blastomyces dermatitidis, Histoplasma capsulatum, Cryptococcus neoformans, Saccharomyces cerevisiae, Ceratostomella ulmi, Typhula gramineum, Sclerotinia sclerotiorum*, and *Endothia parasitica*. Slightly more resistant but inhibited at 3–12 units/ml are *Candida krusei, Trichophyton mentagrophytes*, and *Aspergillus fumigatus*. Many of the oomycetes are relatively resistant; *Saprolegnia ferrax, S. parasitica*, and *Phythophthora parasitica require 100 units/ml for inhibition.*

Resistance of yeasts and fungi to nystatin cannot be developed by serial transfers in vitro through increasing concentrations of the drug (except for occasional slight resistance, up to fourfold). *See* ANTIBIOTIC.

Desirable therapeutic effects are obtained in intestinal moniliasis by oral dosage of 1,500,000 units/day. Vaginal moniliasis is controlled by the use of two vaginal tablets of 100,000 units/day intravaginally. Dusting powder, 100,000 units/g, aids in control of monilial infections of the skin and diaper rash when the causative organism is *Candida albicans* or a similarly sensitive organism

Pharmacology. Nystatin is essentially nontoxic by oral administration. No demonstrable absorption of nystatin follows oral administration Therapeutic activity is exhibited against experimental *Candida* infections in mice by intraperitoneal or subcutaneous administration The drug is not recommended for parenteral administration. No aller*gic reactions are exhibited and no other side reac*tions have been reported.

Production. Commercial production of nystatin is accomplished by fermentation (aerobic growth) with cultures of *Streptomyces noursei* in a manner closely resembling production of other antibiotics. *See* PENICILLIN; TETRACYCLINE.

Several stages of inoculum development, including flasks and inoculum tanks, are used The final fermentation is conducted in tanks of 18,000-gal capacity. Because the organism is strongly aerobic, the medium is aerated with sterile compressed air and mechanically agitated Suitable media may con*tain soybean meal or other high-protein* seed meals, a carbohydrate such as glucose, and calcium carbonate.

Recovery is accomplished by filtration of the mycelium from the broth, followed by solvent extraction of the filter cake. Concentration of the antibiotic by evaporation of the solvent is followed by several purification stages including a final crystallization from a suitable solvent The whole broth may be evaporated and the syrup (or the filter cake noted above) dried for use in animal feeds. *See* MYCOLOGY, MEDICAL.

[R.E.B.]

Bibliography: T H. Sternberg and V. D Newcomer, *Therapy of Fungus Diseases*, 1955; H. Welch and F. Martí-Ibáñez (eds.), *Antibiotics Annual 1954-1955–1957-1958*, 1955-1958.

 *Oak to Ozonization*

ak

genus, *Quercus*, of trees, some of which are ırubby, with about 200 species, mainly in the orthern Hemisphere. About 50 species are native the United States All oaks have scaly winter ıds, usually clustered at the ends of the twigs, ıd single at the nodes The fruit is a nut (acorn) ırrounded at the base by an involucre, the acorn ıp. The pith is star-shaped The leaves are simple ıd usually lobed

Oaks furnish the most important hardwood lum-er in the United States. Principal uses are for harcoal, barrels, building construction, flooring,

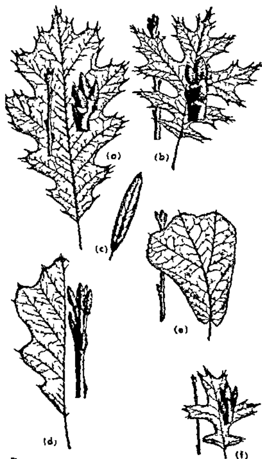


Fig 1 Black oaks (a) Red oak, Quercus rubra (b) Scarlet oak, Q coccinea (c) Willow oak, Q phellos (d) Black oak, Q velutina (e) Blackjack oak, Q marilandica (f) Pin oak, Q palustris.

railroad ties, mine timbers, boxes, crates, vehicle parts, ships, agricultural implements, caskets, woodenware, fence posts, piling, and veneer.

Eastern oaks. The oaks of the eastern United States are divided into two main categories, the black oak group and the white oak group

Black oak group. In this group the leaf lobes are bristle-tipped. Acorns ripen in 2 years, and winter buds are pointed.

The northern red oak, *Q. rubra*, which may attain a height of 75 ft, grows in the eastern half of the United States, except the extreme South and Southeast It can be recognized by the large acorns with flattish cups and by the red, usually shiny, winter buds. In old trees the bark is comparatively smooth with shallow vertical grooves. It is the most important timber tree of the black oak group and is also a popular shade tree.

Other commercially valuable species in the eastern United States include scarlet oak, *Q. coccinea*, a highly prized ornamental tree with deeply cut leaves which turn a brilliant crimson in the fall; pin oak, *Q. palustris*, with tiny acorns and small, deeply cleft leaves, black oak, *Q. velutina*, one of the most common species, with large 5-sided, pubescent (hairy) winter buds and rough, black bark, southern red oak, *Q falcata*, with long, often sickle-shaped leaf lobes, and blackjack oak, *Q marilandica*, with triangular leaves

Eastern oaks with entire or almost entire leaves are the water oak, *Q nigra*, laurel oak, *Q. laurifolia*, and willow oak, *Q phellos*. The lumber of all these species is similar and usually is classed as red oak lumber.

The live oak, *Q. virginiana*, a medium-sized tree of the South Atlantic and Gulf Coast regions, has evergreen leaves.

White oak group. This group has rounded leaf lobes, acorns which ripen in 1 year, and winter buds which are usually rounded (Fig. 2).

White oak, *Q. alba*, furnishes the most valuable hardwood lumber of all Eastern trees. In this same group is the bur oak, *Q macrocarpa*, a large tree of the eastern United States and adjacent Canada. Its acorns are large and edible.

The chestnut oak subgroup is characterized by leaves with numerous small rounded teeth. This subgroup is represented chiefly by the chestnut oak, *Q. prinus*, of the Appalachian Mountain and Ohio Valley regions; the swamp chestnut oak, *Q. michauxi*, and the swamp white oak, *Q. bicolor*.

Western oaks. In the western United States the native oaks do not have the same high commercial

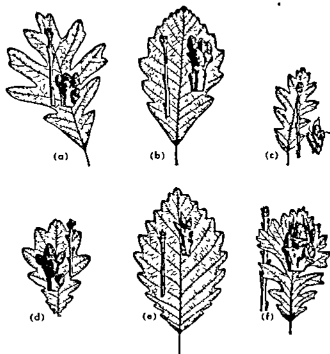


Fig 2. White oaks. (a) White oak, *Quercus alba* (b) Swamp white oak, *Q. bicolor* (c) Turkey oak, *Q cerris*. (d) English oak, *Q robur* (e) Chestnut oak, *Q prinus*. (f) Bur oak, *Q. macrocarpa*. (A H Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

value as timber trees, their place being taken by the valuable western conifers. However, species of importance are the California black oak, *Q. kelloggii*, which grows in Oregon and California; the California live oak, *Q. agrifolia*, with persistent leaves, found in California and Baja California; and the California white oak, *Q. lobata*

The English oak, *Q. robur*, and its varieties are cultivated in the United States, and have leaves of the white oak type, as well as the turkey oak, *Q cerris*, which has shallowly lobed leaves. *See* FOREST AND FORESTRY; TREE. [A H.G.]

Oasis

A fertile spot in a desert where sufficient water is available to support plant and animal life and a permanent human population. Oasis water is obtained in several ways Along rivers such as the Nile and the Indus which arise in rainy lands and flow through deserts, extensive strips of oasis are developed in the valleys Desert water tables are very deep but in some places can be reached by wells dug in the bottoms of dry valleys. Water may reach to or near the surface from artesian systems that are tapped by natural fissures or by drilled wells. In sandy basins the water table may be near the surface and protected by the sand so that natural vegetation or crops will grow in the hollows between the dunes. In mountainous deserts the higher elevations receive more rain; the water descends through the gravels of dry channels and may be recovered by wells in the alluvial fans at the valley mouths.

Oases may be made or improved by deep drilling and power pumps or by transporting water in pipelines. The capacity of an oasis to support people depends directly on the water supply, many oases are densely populated Oasis agriculture is commonly intensive and produces such crops as dates, fruits, cotton, sugar, valuable forages, and grains. In Africa and Asia some oases are centers for the nomadic pastoral activity around them; in the United States some support intensive specialized agriculture, recreational developments or even fairly large cities. [C M.D]

Oats

The grain crop, *Avena* sp., is nearly world wide in distribution. Among the cereal grains, it is exceeded in importance only by corn, rice, and wheat. In the United States every state grows oats, the greatest concentration of acreage being in the eastern half of the country, particularly in the northeastern area. Although the acreage planted has remained fairly constant, uses for this crop have changed with the transition of farm power from horse- to motor-drawn implements. Originally oat grain was used primarily as the preferred feed for horses. Oats now are used in feed mixtures with other concentrates, especially in rations for young animals. In the southern states oats is used as a pasture crop or as a combination pasture and grain crop The 10-year average farm value of oats as a grain crop for the period 1945–1954 was $1,047,501,800 *See* CORN; RICE; WHEAT.

The oats crop requires a relatively cool, moist climate which explains why its best area of adaptation is in the northeast In the southeastern states the crop is fall-sown to utilize the more favorable weather in the fall and early spring. Oats also fits well into row-crop rotation systems as a means of establishing legume-grass forage crops seeded in it

Origin and description. Oats, like most small grain crops, probably originated thousands of years ago Most authorities consider the center of origin to be the Near East or the Mediterranean countries, with China as a secondary center The genus *Avena* is composed of many species which fall into three groups based on 7, 14, and 21 pairs of chromosomes Unlike the genus *Hordeum* (barley), the cultivated species are those with the largest chromosome number (*see* BARLEY). Species in the 7-chromosome group include *A brevis*, a small short-grained species, *A nudibrevis*, a short hull-less-grained species, *A strigosa*, and *A wiestii* Species in the 14-chromosome group are *A. barbata* and *A. abyssinica* In the 21-chromosome group the important cultivated species is *A sativa* Other cultivated species include *A nuda*, a hull-less oat, *A. byzantina*, and *A orientalis*, a side oat, so named because of the location of the seed on the panicle. Noncultivated species in this group are *A fatua*, the common wild oat, and *A sterilis*, the wild red oat. Species within each chromosome number group can be freely intercrossed to produce fertile hybrids. *See* BREEDING (PLANT); REPRODUCTION, PLANT. Hybrids also can be made between the 7- and 14-chromosome group.



Fig 1 Typical panicles of cultivated oats at maturity Note that most of the spikelets produce two grains enclosed within the papery glumes. (J C Allen and Son)

between the 7- and 21-chromosome species are very difficult to produce. Extensive studies have been made to determine the chromosome relationships between species differing in number as a guide to the possible use of interspecific hybrids in developing improved varieties Hybrids between species differing in chromosome number are usually only partially fertile See CHROMOSOME, GENETICS

Within the 21 chromosome cultivated species there are wide variations in types and varieties In all species the inflorescence is a branched panicle with spikelets borne on short pedicles (Fig 1) Spikelets have two to three florets each with staminate and pistillate inflorescences enclosed within the lemma and palea (see GRASS CROPS, INFLORESCENCE) Flowers are normally self-pollinated with occasional outcrossing usually less than 1% Except in hull less varieties of *A nuda*, the lemma and palea adhere loosely to the caryopsis in the mature grain See FRUIT (BOTANY) The oat plant may vary in height from 2 to more than 5 ft and a single plant may produce several culms arising from the lower nodes Among the several varieties the mature grain may have white yellow gray, red, or black glumes Lemmas may be weakly awned or awnless Some oats are spring sown and others are fall sown (see ANNUAL PLANTS) Fall-sown (winter) oats do not possess as high a degree of cold tolerance as winter rye or winter wheat (see RYE).

Varieties. More than 6000 varieties of oats have been listed. These include those in other countries and those developed by breeding programs Among these varieties, there is wide variation in time of maturity, in grain quality, in plant characters, and in resistance to important diseases. Because many plant diseases attack the oat plant, active breeding programs are in progress in the most important oat-producing states to develop new varieties resistant to these diseases. From these breeding programs have come varieties that possess a high degree of resistance to most of the destructive pathogens This has been an important factor in maintaining oat production at its present high level.

Cultural practices. In most states of the Corn Belt, oats are grown after corn in the rotation Corn stalks are disked down or plowed under to prepare the seedbed (see AGRICULTURAL MACHINERY; AGRICULTURAL SOIL AND CROP PRACTICES). Spring oats are seeded as soon as a seedbed can be prepared, usually at the rate of 2½–3 bushels of

bine, either direct from the standing grain or after being windrowed to dry (Fig 2). The moisture content of oats, like other cereal grains, should be not more than 13–14% to avoid heating when stored. Grain used for milling purposes to produce rolled oats for human consumption must be plump and of high quality to command premium prices. See AGRICULTURAL SCIENCE (PLANT); GRAIN CROPS.
[J.J.]

Oat diseases. Like virtually all agricultural crops, oats are plagued with various diseases, major and minor. Stem rust, *Puccinia graminis* var. *avenae*, crown rust *Puccinia coronata* var. *avenae*, two smuts, *Ustilago avenae* and *U kolleri*, and at times root rots, are of major importance, and half a dozen leaf, stem, and kernel blights and several virus infections exact their toll in varying degree, depending on weather and soil conditions. See FUNGI; PLANT VIRUS.

The rusts are the most devastating of all and, in seasons particularly favorable to their development, may reduce the yield to one-half or even one-fourth of normal Both rusts are caused by specific fungus parasites that invade tissues of the oat



Fig 2. Combining an oat crop from the windrow.

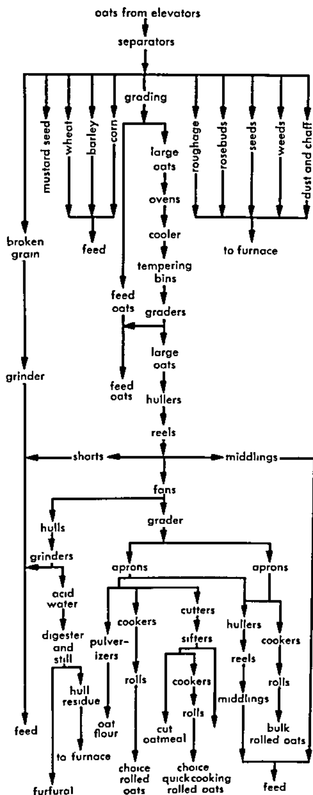


Fig 3 Oat-mill flow sheet. (*From M E. Parker, E. H Harvey, and E. S. Stateler, Elements of Food Engineering, vol 1, Reinhold, 1952*)

plants, where they grow and absorb nutrients from their host. Finally, the fungus bursts through the epidermis of the oat plant in countless places to produce hundreds of thousands of microscopic spores, the fungus "seeds," which invade other oat plants and produce more spores. These rusts can live through the winter in the northern states and start new epidemics in the spring, provided they have the appropriate alternate hosts on which to get started (barberry, *Berberis* sp., for stem rust and buckthorn, *Rhamnus* sp., for crown rust) The most devastating epidemics, however, start in the southern states where the rusts can continue to propagate all through the winter on oats alone In years when abundant rains and dews coincide with the growing seasons over wide areas of the country and when temperatures are favorable, the southern-propagated rust is likely to advance northward over the oat fields as they develop Such waves of destruction may progress from Mexico 2000 miles northward into Canada. Fortunately, the precise combination of circumstances that favor such epidemics occur only occasionally, the last in 1953 and 1954. In other years rust may develop to only a moderate degree or not at all

The loose and covered smuts, *Ustilago avenae* and *U. kolleri*, likewise are fungus parasites However, these invade the oat plants only through the sprout of the germinating seed and only while it is below ground Once inside its host, the parasitic strands of the smut fungus develop unseen, keeping pace with the growth of the oat plant until it heads Then the smut appropriates the host-plant nutrients which are intended to produce grain and changes the young kernels into a mass of black powder (spores), the reproductive bodies of the fungus. These spores contaminate the surface of the grain of near-by healthy plants and there wait to be planted and infect the next crop

Most of the various leaf, stem, and grain spots and root rots are caused by other parasitic fungi and bacteria (*see* BACTERIA). Red leaf, blue dwarf, and mosaic are caused by viruses. Other disorders may be due to nutrient deficiencies, to soil or climatic factors, or to insects (*see* INSECTA; PLANT, MINERALS ESSENTIAL TO).

Disease control. Since oats yield a relatively small return per acre, only inexpensive disease-control practices are economical Treatment of the seed with mercurial fungicides effectively controls smuts and seedling blights (*see* FUNGISTAT AND FUNGICIDE). For rust control, resistant varieties are depended upon which also provide resistance to the smuts and other diseases However, none is resistant to all diseases, and variants in the pathogen sometimes arise that can attack the otherwise resistant varieties Rotation and maintaining a balanced soil fertility are important in the control of root rots and some of the stem and leaf blights *See* FERTILIZING; PLANT DISEASE CONTROL. [M B M]

Processing. The oat kernel has a fibrous hull inedible by humans The goal in milling oats is to obtain the maximum yield of clean uniform, sound, whole oat kernels which are free from hulls, floury material, extraneous matter, and undesirable flavors Oat kernels with the hulls removed are called groats The percentage of hulls can vary from 21 to 43% but average about 25%. An oat-mill process flow diagram is shown in Fig 3

Separating and grading. Oats, at the start of processing, are called green oats and may contain other grains, foreign mat [illegible] d vary [illegible]

ons and gradings according to length, width, density. Length graders consist of shaker eens with various-sized openings. Corn, sticks, l trash are separated from oats on the large-hole ker screens; small seeds, through small-hole eens. Disk separators, which consist of indented tes revolving on a horizontal shaft, are the main e of indent machine Small indents in the rotat- plates lift out seeds and broken material. The t kernel, being long, falls out of the indents on e plate Width graders consist of slotted cylin- ical screen machines which separate thin oats rough narrow slots and milling oats through ider slots Materials such as wheat, double oats, rley, and corn pass through the cylindrical reen Air currents are used to separate light from eavy grain by passing air through the grain as it ills.

Drying Milling oats are subjected to drying or oasting. The usual type of dryer consists of large, pen steel pans, 10-12 ft in diameter and placed ne above another in stacks of 7 to 14 The bottom f each pan is steam-jacketed Oats are fed to the center edge by sweeps as long as the diameter of the pan The sweeps as they ride about on the surface of the pan mix the grain and prevent any oats from remaining too long in contact with the heating surface. When the oats reach the edge of one pan, they fall into chutes which carry them back to the ...

... between 1/2 and 1% It then passes to a cooler where air circulation further reduces the moisture content about another 1%

The roasting process serves several purposes, such as developing flavor improving keeping quality, and facilitating the breaking away of the groats from the hull during milling

Hulling The con...
produce oa...
lar carbor...
the other The lower stone is stationary, the upper one revolves rapidly Dry oats are fed by gravity through an opening in the center of the upper stone and pass between the stones to the ...
distance be...
the space ...
crushing ...
graded for size, including both length and thickness

A relatively recent development in hulling is the use of impact hullers The oats are fed to the center of a high-speed rotor which has a horizontal plate with fins that throw the oats by centrifugal force against a liner The liner may have a covering of special composition rubber. The hulls are loosened from the oat groats by impact One advantage of this type of huller is that the grading is not quite so important as in the rotating stone system, which requires that the space between the stones be adjusted for different sizes of oats.

The following products are obtained from the hullers: hulls, groats, broken groats and meal, flour, unhulled oats, and a small percentage of barley. These materials are separated by air aspiration and screening. The choicest, plumpest groats are used to make package-grade rolled oats, while the less choice groats make either bulk or feed rolled oats. The broken material becomes feed meal.

Products. From milled oats, the following products are produced: steel-cut oats, rolled oats, oat meal, and oat flour. From oat hulls, furfural is manufactured.

Steel-cutting is done with rotary granulators The purpose of cutting is to convert groats to uniform granules. The granules are flaked between rolls to produce a quick-cooking breakfast cereal.

Rolled oats are made by treating the groats with live steam just before rolling. After rolling, the flakes are passed through separators to remove all fine material Rolled oats are used in breakfast foods, cookies, and bread.

Oat flour can be made by several means such as hammer mills, attrition mills, or pulverizers. Oat flours have uses as antioxidants, constituents of baby food, and in soaps and cosmetic preparations

Oat hulls form the starting material for the manufacture of furfural It is made by the destructive distillation of oat hulls in the presence of acid and steam under controlled conditions which change the pentosans in the hulls to pentoses. These are then dehydrated to furfural Furfural is used as a solvent and in the manufacture of plastics and nylon [J.A.SH.]

Bibliography H J. Brounlee and F. L. Gunderson, Oats and oat products. Culture, botany, seed structure, milling, composition, and uses, *Cereal Chem.*, 15:257–272, 1938; M. B Jacobs (ed.), *The Chemistry and Technology of Food and Food Products*, vol. 2, 1944

Obelia

A member of the class Hydrozoa, phylum Coelenterata. Obelia is a small colonial animal, whitish in color and mosslike in appearance It grows attached to rocks, shells, and pilings as deep as 40 fathoms along the Atlantic Coast from Long Island Sound to Labrador, and also along the Pacific Coast of North America

Obelia is a frequently studied animal because its life history is regarded as a classic example of alternation of generations, or metagenesis. The sexually produced generation is asexual and gives rise in turn, to the sexual generation.

The slender branching so-called stem of Obelia is attached to the substrate by special outgrowths resembling roots, the hydrorhiza. The stem is of two layers, an inner, hollow coenosarc with the typical layers of endoderm, mesoglea, ...
derm ...
Th...

cavity and is continuous throughout the colony. The asexual polyps are called hydranths. They are much like those of *Hydra* in structure, being a continuous part of the coenosarc, and are covered with a continuation of the perisarc, called the hydrotheca. Each polyp is armed with about 20 solid tentacles.

The reproductive polyp is called a gonangium. The perisarc is an open-ended vaselike structure called the gonotheca. The perisarc is modified into a central shaft, the blastostyle. The latter bears along its surface several minute, asexually produced medusa buds When mature, the buds escape as small medusae, or jellyfish The sexes of the medusae are separate. Sperm and eggs are shed into the ocean. The zygote develops into a swimming planula which eventually settles to the bottom and transforms into the asexual, colonial, or hydroid, form

Food of both forms is mainly plankton and, as in other Coelenterata, is obtained by the tentacles *See* COELENTERATA; HYDRA; METAGENESIS. [J.D.B.]

Obesity

The presence of an excessive amount or abnormal distribution of fat in an individual. In the United States, there is a certain preoccupation with the normal tendency of the middle-aged person to accumulate more adipose tissue This tendency toward obesity is enhanced by rich diets, lack of regular exercise, and the rejection of certain fundamental health rules. In addition, emphasis has been placed upon the relationship between the overindulgence in food and tension-producing situations. People are said to eat too much because they are anxious, shy, insecure, hostile, or have some other need for which eating can at least partly compensate. *See* NEUROSIS.

Studies in nutrition and metabolism indicate [illegible]

tends to slow down somewhat, more important, however, there are changes in hormonal balance which affect fat utilization by the body. *See* HORMONE.

Although fat was formerly believed to be a relatively inert substance stored in deposits throughout tissues, it has been determined that this fat is one of the most active tissues from a metabolic standpoint. At least four hormones participate in the regulation of fat in relation to carbohydrate-protein metabolism. These include adrenalin, adrenocorticotrophic hormone of the pituitary, thyroxin, and insulin. Others play secondary roles in regulation of body storage, use, and excretion of fat and its products. *See* EPINEPHRINE; HORMONE, ADENOHYPOPHYSEAL; HORMONE, ADRENAL CORTEX; INSULIN.

Aside from the popular overeating theory, and the newer but still incomplete understanding of fat metabolism, obesity occasionally appears in relation to specific disease processes.

These include certain disorders of the hypothalamus or injury to it resulting from changes in adjacent structures, such as in the formation of some pituitary tumors Froehlich's syndrome, characterized by a feminine pattern of obesity and sexual dysfunction, is thought to result from associated damage to the neighboring hypothalamus, although it was formerly believed to be the result of hypopituitarism Similar symptoms can be induced in animals when the hypothalamus is injured.

In some types of adrenal dysfunction, such as Cushing's syndrome and hyperadrenocorticism, a peculiar obesity of the neck, trunk, and shoulders is typical. In certain cases of diabetes mellitus obesity is frequently seen at some time during the course of the disease. An uncommon type of obesity in which the excess fat deposits are tender and painful is known as adiposa dolorosa.

The influence of hereditary factors cannot be overlooked because it has been established that obesity and other body characteristics tend to be transmitted from one generation to the next.

Persistent excessive obesity should be medically evaluated, and no rigorous dieting or exercising should be undertaken without medical supervision. [E.G.ST.]

Observatory, astronomical

A building or group of buildings which house fixed astronomical instruments, generally including one or more equatorial telescopes for the examination of celestial objects.

Small observatories. Amateur observatories are usually built, equipped, and operated by their owners. They often contain a reflecting telescope

Academic observatories in America are found at most state universities and colleges where there are departments of astronomy The instruments may be a refractor, rarely exceeding 12 in aperture, a transit, clocks, and some portable instruments. Telescope attachments may include a filar micrometer, visual photometer, solar eyepiece, solar spectroscope, and means of photographing the Sun and Moon *See* ASTRONOMICAL PHOTOGRAPHY; ASTRONOMICAL SPECTROSCOPY

Public observatories have telescopes, usually refractors, connected with planetariums, to permit visitors to see celestial objects currently visible in the evening sky (*see* PLANETARIUM). The large professional telescopes made for astronomical research are ill-adapted to public use on account of their size and the hazards of their operation As with any large machines, insurance must be carried for all operators

Large observatories. Great professional observatories are usually the gift of some man of large fortune The investment is handled by a trust fund or institution established by the donor and is dedicated to the accumulation of human knowledge Often a university administers the fund and the observatory. Present-day interest in space around us has caused governments to establish observatories for special purpose



The 200-in observatory on Mount Palomar, Calif at night Photograph by moonlight

Large telescopes are principally placed on high levations which have high percentage of clear sky ith minimum atmospheric optical instability or bad seeing"

Professional observatories include the largest elescopes in the world Those erected in the nineeenth century generally have refractors The larg est refractors are the 40-in at the Yerkes Observatory, Williams Bay, Wis 1897, the 36-in at the Lick Observatory Mount Hamilton, Calif., 1888; and the 33-in at Meudon, France, 1889 The problems of astronomy were then solved by visual observations.

When dry plates of high speed and uniform quality became available about 1890 the trend toward astronomical photography began Because the reflector is perfectly achromatic it is well adapted to photography Thus all the newer large telescopes under construction or in contemplation are reflectors The largest reflectors are the 200-in at Mount Palomar 1948, the 120 in at the Lick Observatory, 1959, and the 100-in at Mount Wilson, 1918 all in California Reflectors can be used at three focal lengths for direct photography, photoelectric and thermoelectric measurements, spectroscopy and other special procedures Schmidt reflectors are an exception because a lens is combined with a mirror in them (see TELESCOPE, ASTRONOMICAL)

A telescope is housed in a building with rotatable dome A slit with movable cover enables the telescope to be exposed to the sky in any direction as illustrated Domes of conical form or frustrums of cones and even sliding roofs are used

Special-purpose observatories. Many special-purpose observatories are built. National observatories, such as the U.S Naval Observatory, are primarily dedicated to checking the data annually issued in an astronomical ephemeris used by astronomers, navigators, and surveyors They also provide radio time signals from electronic clocks checked by an observatory equipped with a photographic zenith tube accurate to about 0 001 seconds. Through the cooperation of several observatories in widely different longitudes, the position of the Earth's poles can be measured by the use of photographic zenith tubes.

Photographic observatories may use large camera lenses, mounted as equatorial telescopes, to chart the heavens. Lenses of aperture as great as 24 in. are used. The Allegheny Observatory at Pittsburgh, Pa., has a 30-in telescope with objective corrected to photographic light. Schmidt telescopes with spherical mirrors and correcting plates at the center of curvature are also used The Big Schmidt at Mount Palomar has a 72-in. mirror with 48-in. correcting plate.

Solar observatories may use simple-lens equatorial refractors with narrow-band filters or coelostats with fixed vertical telescopes in tower form. See SUN.

Radio observatories Radio astronomy observatories have an external telescope consisting of a large paraboloidal frame covered with metal to reflect the celestial radio waves to a small antenna at the focus. The signals are wired to a neighboring laboratory. The largest such radio telescope—of 250-ft aperture—is at Jodrell Bank, near Manchester, England Another, of 140-ft aperture, is being placed in operation at Green Bank, W. Va. See RADIO ASTRONOMY

Space observatories Space telescope observatories are being considered A 50-in reflector is being set up on Kitt Peak in Arizona to be operated from Tucson, 45 miles away by electronic control. When proper performance is secured, one of these telescopes will be orbited into space 25,000 miles high, moving eastward in an equatorial orbit, where it will appear to be stationary in the sky. It will be operated from Tucson See ASTRONOMICAL INSTRUMENTS, CHRONOMETER; CORONOGRAPH; RADIO TELESCOPE, SPECTROHELIOSCOPE.

[E P.]

Bibliography· H. C. King, *The History of the Telescope*, 1955.

Obsessive compulsive reaction

A type of neurosis The characteristic symptom pattern of the obsessive compulsive is the irrational persistence of ideation and enactment which is recognized as irrational but which the person cannot alter

and f

may h

case of obsessive symptoms and, in compulsive behavior, the person is driven to perform acts whose relevance for his problems remains utterly obscure

cavity and is continuous throughout the colony. The asexual polyps are called hydranths. They are much like those of *Hydra* in structure, being a continuous part of the coenosarc, and are covered with a continuation of the perisarc, called the hydrotheca. Each polyp is armed with about 20 solid tentacles.

The reproductive polyp is called a gonangium. The perisarc is an open-ended vaselike structure called the gonotheca. The perisarc is modified into a central shaft, the blastostyle. The latter bears along its surface several minute, asexually produced medusa buds. When mature, the buds escape as small medusae, or jellyfish. The sexes of the medusae are separate. Sperm and eggs are shed into the ocean. The zygote develops into a swimming planula which eventually settles to the bottom and transforms into the asexual, colonial, or hydroid, form.

Food of both forms is mainly plankton and, as in other Coelenterata, is obtained by the tentacles. *See* COELENTERATA; HYDRA; METAGENESIS [J.D.B.]

Obesity

The presence of an excessive amount or abnormal distribution of fat in an individual. In the United States, there is a certain preoccupation with the normal tendency of the middle-aged person to accumulate more adipose tissue. This tendency toward obesity is enhanced by rich diets, lack of regular exercise, and the rejection of certain fundamental health rules In addition, emphasis has been placed upon the relationship between the overindulgence in food and tension-producing situations. People are said to eat too much because they are anxious, shy, insecure, hostile, or have some other need for which eating can at least partly compensate. *See* NEUROSIS

Studies in nutrition and metabolism indicate that although such factors certainly may play a role in many cases, more subtle differences occur among individuals. As a person ages, his metabolic rate tends to slow down somewhat, more important, however, there are changes in hormonal balance which affect fat utilization by the body. *See* HORMONE

Although fat was formerly believed to be a relatively inert substance stored in deposits throughout tissues, it has been determined that this fat is one of the most active tissues from a metabolic standpoint At least four hormones participate in the regulation of fat in relation to carbohydrate-protein metabolism. These include adrenalin, adrenocorticotrophic hormone of the pituitary, thyroxin, and insulin. Others play secondary roles in regulation of body storage, use, and excretion of fat and its products. *See* EPINEPHRINE; HORMONE, ADENOHYPOPHYSEAL; HORMONE, ADRENAL CORTEX; INSULIN.

Aside from the popular overeating theory, and the newer but still incomplete understanding of fat metabolism, obesity occasionally appears in relation to specific disease processes

These include certain disorders of the hypothalamus or injury to it resulting from changes in adjacent structures, such as in the formation of some pituitary tumors. Froehlich's syndrome, characterized by a feminine pattern of obesity and sexual dysfunction, is thought to result from associated damage to the neighboring hypothalamus, although it was formerly believed to be the result of hypopituitarism. Similar symptoms can be induced in animals when the hypothalamus is injured

In some types of adrenal dysfunction, such as Cushing's syndrome and hyperadrenocorticism, a peculiar obesity of the neck, trunk, and shoulders is typical. In certain cases of diabetes mellitus obesity is frequently seen at some time during the course of the disease. An uncommon type of obesity in which the excess fat deposits are tender and painful is known as adiposa dolorosa.

The influence of hereditary factors cannot be overlooked because it has been established that obesity and other body characteristics tend to be transmitted from one generation to the next.

Persistent excessive obesity should be medically evaluated, and no rigorous dieting or exercising should be undertaken without medical supervision. [E.G.S.]

Observatory, astronomical

A building or group of buildings which house fixed astronomical instruments, generally including one or more equatorial telescopes for the examination of celestial objects.

Small observatories. Amateur observatories are usually built, equipped, and operated by their owners. They often contain a reflecting telescope

Academic observatories in America are found at most state universities and colleges where there are departments of astronomy The instruments may be a refractor, rarely exceeding 12 in aperture, a transit, clocks, and some portable instruments Telescope attachments may include a filar micrometer visual photometer, solar eyepiece, solar spectroscope, and means of photographing the Sun and Moon *See* ASTRONOMICAL PHOTOGRAPHY, ASTRONOMICAL SPECTROSCOPY.

Public observatories have telescopes, usually refractors, connected with planetariums, to permit visitors to see celestial objects currently visible in the evening sky (*see* PLANETARIUM). The large professional telescopes made for astronomical research are ill-adapted to public use on account of their size and the hazards of their operation As with any large machines, insurance must be carried for all operators

Large observatories. Great professional observatories are usually the gift of some man of large fortune The investment is handled by a trust fund or institution established by the donor and is dedicated to the accumulation of human knowledge Often a university administers the fund and the observatory. Present-day interest in space around us has caused governments to establish observatories for special purpo

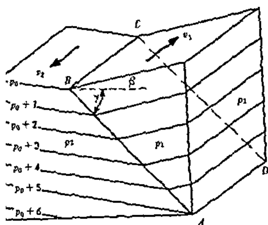


Fig 1 Block diagram showing slope of isobaric surface p_0, p_0+1, p_0+2, . . . , with slope angle β, discontinuity surface $ABCD$ with slope angle γ, in two homogeneous water masses of density ρ_1 and ρ_2; currents with velocities v_1 and v_2 in Northern Hemisphere

press themselves in a slope of the surface of the sea and a slope of the surfaces of equal pressure perpendicular to the direction of current In Fig. 1, p_0, p_0+1, and p_0+2 indicate such surfaces with the slope angle β These slopes are as a rule very mall and cannot be seen or directly measured. ven in such strong currents as the Gulf Stream he slope of the sea surface achieves a gradient of only 10 cm over a distance of 10 km In the sea here are frequently layers of varying density which separate specifically light from specifically heavy water masses, these discontinuity layers also tend to be inclined perpendicular to the direction of the current Such a discontinuity layer is illustrated in Fig 1 by $ABCD$ and the angle of inclination γ In the Northern Hemisphere the surface of the sea rises from left to right when one is looking in the direction of the current. The slope of the discontinuity layer is about a thousand times greater than that of the surface of the sea This may be shown by a cross section through the Gulf Stream from the Chesapeake Bay to Bermuda (Fig 2) There the discontinuity layer reaches the surface of the sea and two masses of water appear with a sharp boundary between them The Cold Wall in the Gulf Stream represents such a division.

Dynamic computations. A quantitative determination of the ocean currents can be made by calculation of the varying vertical distances of the isobaric surfaces from the vertical density distribution, as observed at a representative number of oceanographic stations The differences in velocity between two levels can be obtained from the relationship, first given by B Helland Hansen and J. W. Sandstrom,

$$v_1 - v_2 = 10(D_A - D_B)/fl$$

In this relationship $v_1 - v_2$ is the difference in velocity, $D_A - D_B$ is the difference in level between selected isobaric surfaces between the two stations A and B, f is the Coriolis acceleration, and l is the distance between A and B. This relationship forms the most important basis for all statements regarding the distribution of currents in the oceans and seas of the world. The method, however, provides only differences of velocity. A streamless layer must be indirectly obtained from the stratification as a zero layer (layer of no motion) if absolute values for velocity and for volume transport of water are to be determined, as for instance, in the example of a section through the Gulf Stream shown in Fig. 2. The Gulf Stream is a relatively narrow band of water (approximately 40 km wide) but relatively deep (about 1000 m) Current speeds in the Gulf Stream reach 150 cm/sec (~3 knots). The volume transport also is enormous, 87×10^6 m³/sec, a figure which is 65 times greater than the water transport of the rivers of all the continents.

Wind stress. Air does not merely glide along the surface of the water but exercises a frictional effect, or wind stress, which causes the surface water to be carried along with it. The movement of this thin layer on the surface of the water is conveyed by an internal turbulent friction to the deeper levels The eventual result of such interaction, in a limitless homogeneous sea under the influence of a wind which remains steady, would be a pure drift current, the theory of which was developed by V. W. Ekman in 1902. The resulting current distribution is illustrated by the so-called Ekman spiral (Fig. 3). According to the theory, there is a water movement on the sea surface of the Northern Hemisphere at an angle of 45° to the right of the direction of the wind and at a speed approximately 1.5% of the wind speed. With increasing depth the current turns farther towards the right and gradually subsides. When the direction of this current reaches an angle of 180° to the flow on the sea surface, the speed of the current is only 1/23 that of the surface water. This depth is called the depth of frictional influence For example, at a latitude of 50° this depth amounts to 60 m when the wind speed reaches 7 m/sec (13.8 knots) The resultant volume transport of the pure drift current, when taken over the whole layer between the sea surface and the depth of frictional influence, runs perpendicular to the direction of the wind turned in a clockwise sense It is understandable, therefore, that upwelling of cold deep water occurs at a coast when the wind blows parallel to the coast, with the coast on the left-hand side of the wind [illegible]

[illegible] pressure force [illegible] are [illegible] by th[illegible] then [illegible] and with it the slope of the surfaces of equal density. These alterations in slope are accompanied by cross circulations, to the direction of the main current, which contain vertical current components. They are very small indeed; for instance in the zone of upwelling along the coast of California the vertical velocity is only 0.003 cm/sec and represents also only 1 part per

to him, for example, repeated handwashing, sometimes to a self-injurious point.

The psychological gain that derives from these symptoms is an isolation and stabilization of fragments drawn from a broader range of problems with which the person has been unable to cope, the seeming object being to handle the unrecognized broader problems by excessive concern and preoccupation with the fragments. Thus, a rigid ritual, handwashing, may serve to concretize guilt feelings which it does not, however, succeed in resolving. One also observes symptoms that express to an exaggerated degree fragments of an unconscious wish, as in the constricting obsessive thought that one will kill some particular person toward whom one consciously has no hostile feelings but toward whom one may have deeply repressed or displaced unconscious hatred.

Whatever the symptom pattern in neurosis, there seems to be a certain number of common dynamic features to be observed In all cases they appear to involve unrecognized and unacceptable impulses that must find indirect expression Similarly, they all involve relatively inadequate but costly defenses for denaturing or getting rid of these impulses, the very inadequacy of the defense leading to the arousal of anxiety which either directly or by conversion has a further disrupting effect on behavior. The costliness of the neurotic's failing defenses is inherent in his tendency to adopt new forms of self-defense that preoccupy him to such a degree that he is not able to cope with either the threatening inner impulses or the realities of his life situation. *See* ABNORMAL BEHAVIOR, NEUROSIS [J.S.B., W.MS]

Obsidian

A volcanic glass, usually of rhyolitic composition, formed by rapid cooling of viscous lava. The color is jet-black because of abundant microscopic, embryonic crystal growths (crystallites) which make the glass opaque except on thin edges. Iron oxide dust may produce red or brown obsidian.

Obsidian usually forms the upper parts of lava flows. Well-known occurrences are Obsidian Cliffs in Yellowstone Park, Wyoming; Mt. Hekla, Iceland; and the Lipari Islands, off the coast of Italy. Less commonly, obsidian forms selvages of dikes and sills. *See* IGNEOUS ROCKS; VOLCANIC GLASS [C.A.CA]

Occultation

The apparent disappearance of a star or planet behind the surface of the Moon, or of a satellite behind the disk of the parent planet.

Occultations by the Moon are observed to determine the position of the Moon at the time of the phenomenon. They are also used for precise determinations of longitude. Because of the eastward motion of the Moon against the background of stars, immersion (or disappearance) of the star or planet occurs at the eastern limb, and emersion (or reappearance) at the western limb. A single occultation is visible only from a certain region of Earth. The parallels of latitude which enclose the region of visibility are called limiting parallels. Outside that region, the effect of parallax causes the star or planet to remain clear of the lunar disk. Location of the observing site within the limiting parallels is a necessary, but not sufficient, condition of visibility. It is also necessary that the Moon be above the horizon at the time of the occultation for the location considered.

Occultations of the four Galilean satellites of Jupiter occur at each of their superior conjunctions, with the exception of satellite IV, which occasionally passes clear of the planet's disk. In general, the immersions cannot be observed when Jupiter's shadow extends westward from the planet. Similarly, the emersions cannot be observed when the shadow extends eastward. In the first instance, the satellite is still eclipsed when the occultation begins, whereas in the latter, the satellite is in eclipse as the occultation ends. *See* ECLIPSE, ASTRONOMICAL; JUPITER. [S.D.C.]

Ocean currents

All dislocations of water masses within the ocean, including the permanent current systems which are a part of the general circulation; horizontal water movement in the surface and the deeper layers, vertical water movements, such as upwelling and sinking; and variable currents as caused by tides and large-scale turbulence. Among the important aspects of ocean currents are their speed and direction, the mass (or volume) transport, the rates at which mixing takes place, the variations of these parameters with time, and the mechanisms which produce and maintain the circulation of the oceans.

This discussion treats the dynamics of ocean currents, characteristics of the surface and deep circulation, and methods of measuring currents and mixing For a discussion of water movements within estuaries, also the distribution of physical properties in sea water, the characteristics of water masses, and mixing and stirring processes see ESTUARINE OCEANOGRAPHY; SEA WATER

DYNAMICS OF OCEAN CURRENTS

Ocean currents are the result of different forces: primary forces, which give rise to currents and maintain them, and secondary forces, which influence motions already existing The primary forces include internal and external pressure forces, the stress of the wind, and the tide-generating forces. Secondary forces are the frictional force and the deflecting force of the rotation of the earth, which is also called the Coriolis force *See* CORIOLIS ACCELERATION AND FORCE; TIDE.

Pressure forces. Internal pressure forces are based on differences in density, which are maintained by horizontal differences in temperature and salinity as well as by horizontal pressure differences caused by the wind-produced piling-up of water External pressure forces are based on alterations in atmospheric pressure. The pressure fo ex

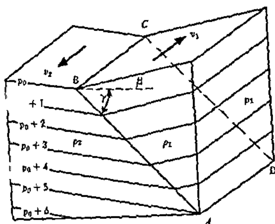


Fig 1 Block diagram showing slope of isobaric surface p_0, $p_0 + 1$, $p_0 + 2$, . . , with slope angle β; discontinuity surface $ABCD$ with slope angle γ, in two homogeneous water masses of density ρ_1 and ρ_2; currents with velocities v_1 and v_2 in Northern Hemisphere.

press themselves in a slope of the surface of the sea and a slope of the surfaces of *equal pressure* perpendicular to the direction of current. In Fig 1, p_0, $p_0 + 1$, and $p_0 + 2$ indicate such surfaces with the slope angle β These slopes are as a rule very small and cannot be seen or directly measured Even in such strong currents as the Gulf Stream the slope of the sea surface achieves a gradient of only 10 cm over a distance of 10 km In the sea there are frequently layers of varying density which separate specifically light from specifically heavy water masses, these discontinuity layers also tend to be inclined perpendicular to the direction of the current Such a discontinuity layer is illustrated in Fig 1 by $ABCD$ and the angle of inclination γ In the Northern Hemisphere the surface of the sea rises from left to right when one is looking in the direction of the current The slope of the discontinuity layer is about a thousand times greater than that of the surface of the sea This may be shown by a cross section through the Gulf Stream from the Chesapeake Bay to Bermuda (Fig 2) There the discontinuity layer reaches the surface of the sea and two masses of water appear with a sharp boundary between them The Cold Wall in the Gulf Stream represents such a division

Dynamic computations. A quantitative determination of the ocean currents can be made by calculation of the varying vertical distances of the isobaric surfaces from the vertical density distribution, as observed at a representative number of oceanographic stations The differences in velocity between two levels can be obtained from the relationship first given by B Helland-Hansen and J W Sandstrom,

$$v_1 - v_2 = 10(D_A - D_B)/fl$$

In this relationship $v_1 - v_2$ is the difference in velocity, $D_A - D_B$ is the difference in level between selected isobaric surfaces between the two stations A and B, f is the Coriolis acceleration, and l is the distance between A and B. This relationship forms the most important basis for all statements regarding the distribution of currents in the oceans and seas of the world. The method, however, provides only differences of velocity. A streamless layer must be indirectly obtained from the stratification as a zero layer (layer of no motion) if absolute values for velocity and for volume transport of water are to be determined, as for instance, in the example of a section through the Gulf Stream shown in Fig. 2. The Gulf Stream is a relatively narrow band of water (approximately 40 km wide) but relatively deep (about 1000 m). Current speeds in the Gulf Stream reach 150 cm/sec (~3 knots). The volume-transport also is enormous, 57×10^6 m^3/sec, a figure which is 65 times greater than the water transport of the rivers of all the continents.

Wind stress. Air does not merely glide along the surface of the water but exercises a frictional effect, or wind stress, which causes the surface water to be carried along with it The movement of this thin layer on the surface of the water is conveyed by an internal turbulent friction to the deeper levels The eventual result of such interaction, in a limitless homogeneous sea under the influence of a wind which remains steady, would be a pure drift current, the theory of which was developed by V. W. Ekman in 1902. The resulting current distribution is illustrated by the so-called Ekman spiral (Fig 3) Accordin[illegible] ment, [illegible] at an [illegible] the w[illegible] at a speed approximately 1.5% of the wind speed With increasing depth the current turns farther towards the right and gradually subsides When the direction of this current reaches an angle of 180° to the flow on the sea surface, the speed of the current is only 1⁄23 that of the surface water. This depth is called the depth of frictional influence For example, at a latitude of 50° this depth amounts to 60 m when the wind speed reaches 7 m/sec (13.8 knots). The resultant volume transport of the pure drift current, when taken over the whole layer between the sea surface and the depth of frictional influence, runs perpendicular to the direction of the wind turned in a clockwise sense. It is understandable [illegible] [illegible] the wind, or even [illegible] wind [illegible] e[illegible] pressure forces which are conditioned by the distribution of density and by the wind-driven currents. Should the wind alter, then the current also alters and with it the slope of the surfaces of equal density These alterations in slope are accompanied by cross circulations, to the direction of the main current, which contain vertical current components. They are very small indeed, for instance in the zone of upwelling along the coast of California the vertical velocity is only 0.003 cm/sec and represents also only 1 part per

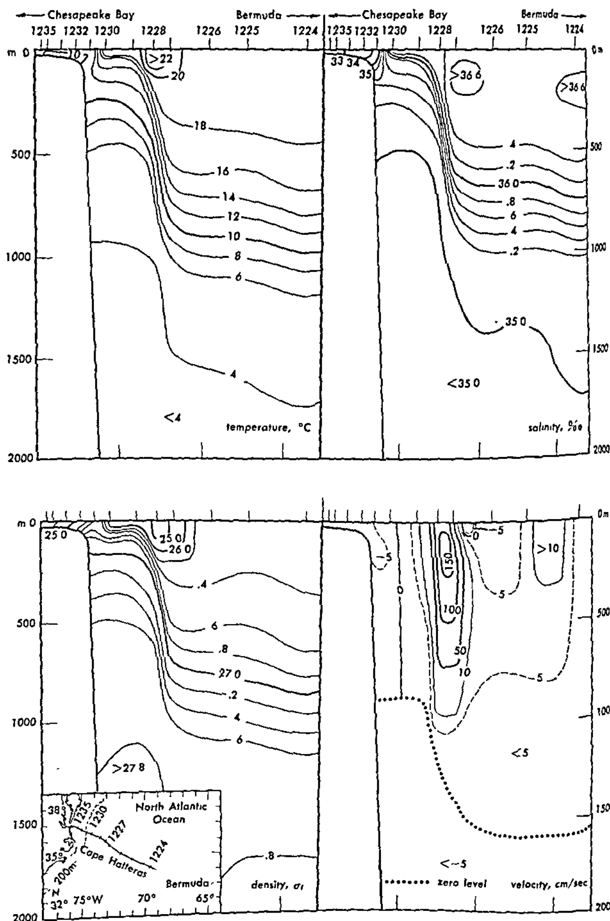


Fig. 2. Cross section of Gulf Stream showing vertical distribution of temperature, salinity, density, and velocity. Current flow in southwest-northeast direction (After G. Dietrich, 1957, based on section by RV lantis, Woods Hole Oceanographic Institution, Stati 1224–1235)

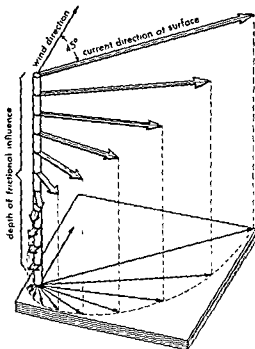


Fig. 3 A schematic representation of a pure wind current in deep water, showing *decrease in velocity* and change of direction at regular intervals of depth (the Ekman spiral)

thousand of the horizontal current components However, because the vertical differences of temperatures and nutrient salts in sea water are much greater than the horizontal differences, this upwelling of water (which is equivalent to 80 m/month) has an important effect. It is sufficient to explain the low temperatures and also the high values of nutrient salts which are associated with great organic productivity in the surface layer of the regions of upwelling *See* SEA WATER FERTILITY

In the latest theory of ocean currents apart from cross circulation, the dependence of the Coriolis forces on the geographical latitude has been taken into consideration by H Stommel and the effect of lateral mixing by C G Rossby Stommel explains why currents on the western sides of the oceans, which are directed towards the pole, are narrow and strong as in the Gulf Stream and in the Kuroshio (Fig 4) Some of the details about the course of a current such as counter currents at the side of and beneath strong currents and meanders in strong currents, their continuation in the direction of the current, and their dissolution in the form of cyclonic eddies of the main current, may be explained theoretically *See* GULF STREAM

SURFACE AND DEEP CIRCULATION

Surface currents. The system of surface currents is restricted mainly to the upper 100–200 m of the sea In the lower geographical latitudes it is shallower and in the higher latitudes it generally is deeper. The speeds of the surface currents remain mostly below 20 cm/sec (0 4 knot). Exceptions to this are found in the Gulf Stream, the Kuroshio, the Agulhas Current, and the Equatorial Counter Currents of the three oceans, all of which have measurable velocities of 1–2 m/sec (2–4 knots). Knowledge of the surface currents is based on measurements of the current and more particularly on dead reckoning from ships. *See* DEAD RECKONING.

The causes of surface currents are wind stress and internal pressure forces. Frictional and Coriolis forces influence the surface currents. The effect of the wind is at its greatest when the direction and strength of the wind are steady; this is so for the lower and middle geographical latitudes. In these latitudes an anticyclonic current system corresponds to the anticyclonic wind system (Fig. 5). Surface currents which flow in a westerly direction in the lower latitudes are parts of this system (the North and South Equatorial Currents of the three oceans). The continuation of these currents is found along the eastern sides of the continents in narrow and strong surface currents directed towards the poles (western boundary currents), for example, the Gulf Stream, Brazil Current, Somali Current (only in the summer of the Northern Hem[illegible]
sphere [illegible]
tralia ([illegible]
these currents turn and flow in an easterly direction

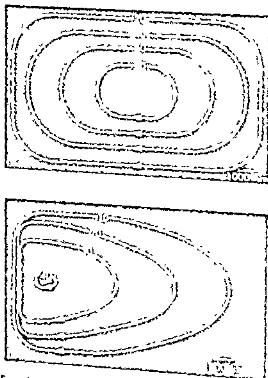


Fig 4 Streamlines showing currents in the case of (a) an ocean on a nonrotating globe, (b) an ocean on a uniformly rotating globe in which the Coriolis forces increase with the geographic latitude (After H Stommel, *The Gulf Stream*, Univ. of California Press, 1958)

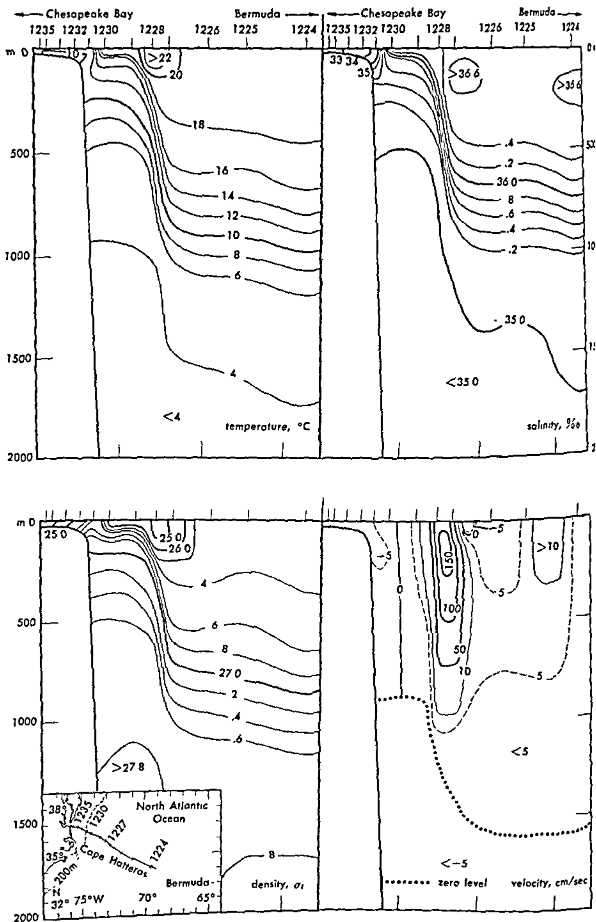


Fig 2 Cross section of Gulf Stream showing vertical distribution of temperature, salinity, density, and velocity. Current flow in southwest-northeast direction (After G. Dietrich, 1957, based on section by lantis, Woods Hole Oceanographic Institution, S 1224–1235)

(North Atlantic Current, North Pacific Current, and West Wind Drift of the Southern Hemisphere). On the eastern sides of the oceans this circulation pattern is completed by surface currents which are directed towards the Equator (eastern boundary currents), for example, the Canaries Current, Benguela Current, West Australia Current, California Current, and Humboldt Current Inbedded in the system of the North and South Equatorial Currents are the Equatorial Counter Currents which are found in the neighborhood of the Equator in all three oceans

In the temperate and higher geographical latitudes, where variable winds are to be found, the internal pressure forces are predominant in their influence on surface currents These currents tend to follow the coasts and shelf edges and, in the Northern Hemisphere, in a manner so that the continents lie to the right-hand side of the current when one looks in the downstream direction (Fig. 5) Examples are the Norwegian, East Greenland, West Greenland, Labrador, and Alaska currents Islands are in this fashion, so to speak, surrounded by currents moving in a clockwise direction (for example, Iceland by the Irminger, North Iceland and East Iceland currents) It is for this reason that the western sides of continents in these geographical latitudes are bordered by comparatively warm waters coming from lower geographical latitudes, whereas off the eastern coasts in the same latitudes there are cold waters from higher geographical latitudes For example, in the Norwegian Current the surface temperature in summer is 10C° higher than in the East Greenland Current; both are in the same geographical latitude. Thus the surface currents are of very great climatic importance All surface currents contain vertical components which vary from [illegible]

[illegible] The vertical components are certainly very small (for example, a speed of 0 003 cm/sec in the upwelling area of the California Current), but in the long run they are of great importance when considering the balance of heat and all sea-contained substances. See MARINE INFLUENCE ON WEATHER AND CLIMATE.

Deep circulation. The deep circulation results in part from the wind stress and in part from the internal pressure forces which are maintained by the budget of heat and salt of the water. Both groups of forces are dependent upon atmospheric influences. Apart from Coriolis and frictional forces the topography of the sea bottom exercises a decisive influence on the course of deep circulation.

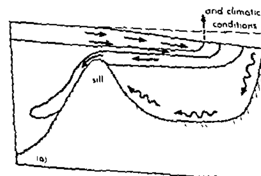
Deep circulation of marginal seas The deep circulation in marginal seas depends largely on the climate of the region, whether arid or humid.

1 Arid climates Under the influence of an arid climate the evaporation of water is greater than the precipitation The marginal sea is therefore filled with relatively salty water of a high density, and its surface lies at a lower level than those of the neighboring ocean Examples of this type are the [illegible]

At the sill depth between two seas, that is, the greatest depth at which there is free horizontal communication, there is to be found water of a low density from the ocean. The water from the marginal sea therefore flows over the sill into the ocean, where it sinks to a level in which it finds other water corresponding to its density At this level it then spreads out horizontally The waters from the Mediterranean Sea and from the Red Sea, because of their high salinity, can be followed far out into the Atlantic and Indian Oceans respectively In the upper layer the oceanic water flows into the marginal seas following the line of slope. See INDIAN OCEAN, MEDITERRANEAN SEA.

2. Humid climates. The deep circulation of marginal seas in humid climates shows a completely different pattern, however, shown schematically in

Atlantic Ocean — Strait of Gibraltar — Mediterranean Sea
Indian Ocean — Bab el Mandeb — Red Sea
Indian Ocean — Strait of Hormuz — Persian Gulf



Atlantic Ocean — Fiord rises — fiords of Norway and Greenland
Aegean Sea — Bosporus — Black Sea
Kattegat — Darsser and Drogden rise — Baltic Sea

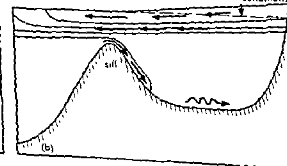
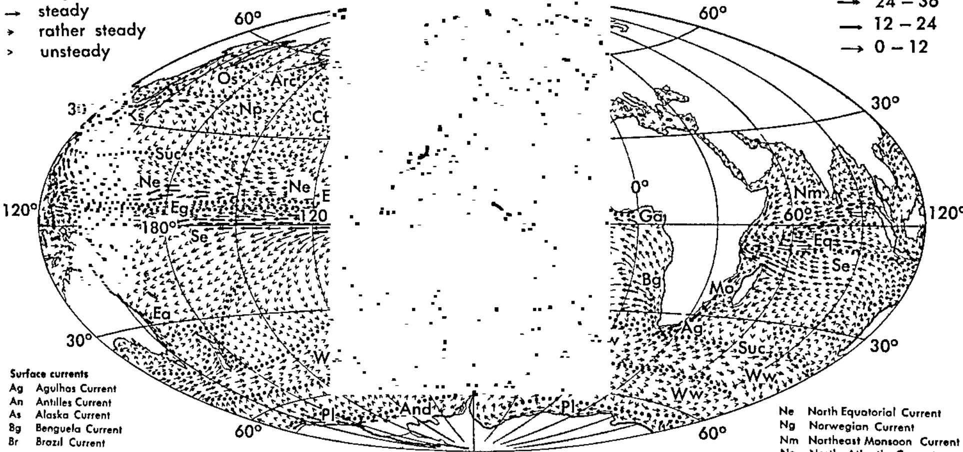


Fig 6 Schematic representation of the circulation in marginal seas and over their sills (a) in arid climates, for example, Mediterranean Sea, Red Sea, and Persian Gulf, (b) in humid climates, for example, Black Sea, Baltic Sea, and fiords of Norway and Greenland.



Surface currents

Ag Agulhas Current
An Antilles Current
As Alaska Current
Bg Benguela Current
Br Brazil Current
Ca Canaries Current
Cf California Current
Ch Cape Horn Current
Ea East Australia Current
Eg East Greenland Current
Eq Equatorial Countercurrent
Fa Falkland Current
Fl Florida Current
Ga Guinea Current
Gu Gulf Stream
Hu Humboldt Current
Ir Irminger Current
Ks Kuroshio
La Labrador Current
Mo Mozambique Current
Ne North Equatorial Current
Ng Norwegian Current
Nm Northeast Monsoon Current
No North Atlantic Current
Np North Pacific Current
Os Oyashio
Pl Polar Current Wind Drift
Po Portugal Current
Se South Equatorial Current
Wd Weddell Current
Ww West Wind Drift

Current limits:

Anc Antarctic Convergence
And Antarctic Divergence
Arc Arctic Convergence
Suc Subtropical Convergence

(After G. Schott, 1943)

1 *The Norwegian Sea*, where the water with the highest density of the world oceans is formed This supplies the North Polar Sea with cold deep and bottom water as well as the North Atlantic Ocean with cold bottom water to a latitude of approximately 50°N The distribution of these waters was the object of investigations in the International Geophysical Year 1958 (made by about 30 research ships in the Polarfront Survey).

2 The Antarctic continental slope in the Weddell Sea, where water temperatures beneath the winter pack-ice are as low as −1.9°C. Favored by the topography of the ocean floor, this water spreads out in the West Atlantic Trough northward to the foot of the Grand Banks (Newfoundland *Basin*) as well as through the Romanche Deep on the Equator into the East Atlantic Trough It also spreads northward into the bottom water of the Indian and Pacific oceans

3. In the Labrador and Irminger seas where the Deep Water (between depths of 1000 and 4000 m) of the Atlantic Ocean originates. Apart from low temperatures this water is distinguished by its richness in oxygen

4 The polar front at a latitude of 50°S Here cold water with low salinity is formed and feeds the subantarctic intermediate water of the Atlantic, Indian, and Pacific oceans.

5. The polar front in the North Pacific Ocean, where subarctic intermediate water is formed Distribution of this water into the North Pacific Ocean is restricted

The deep-sea circulation of the Atlantic Ocean is by far the most active in comparison with that of the Indian and Pacific Oceans because the most important centers of thermohaline convection are found in the Atlantic In addition, the continental barrier of South America forces large masses of water from the surface currents of the South Atlantic into the North Atlantic Ocean In order to compensate for the loss of surface water in the South Atlantic there is a more active deep-water circulation, in which North Atlantic Deep Water flows southward into the South Atlantic Ocean

Tidal streams. The periodic variations in the level of sea waters—the tides of the sea—are always accompanied by periodic horizontal movements of water called tidal streams The tides of the sea are long waves and therefore the tidal streams affect all the water of the oceans from the surface down to the bottom. The tidal stream which runs at flood is called the flood stream, and that which runs at ebb the ebb stream In the vicinity of land the flood and ebb streams run parallel to the coast as rectilinear or alternating currents Between the times of flood and ebb there are a few minutes when there is no stream whatsoever This period is called slack water Following slack water the current flows in the opposite direction. In the open sea the tides are seldom alternating; they turn either clockwise or counterclockwise and are rotating streams In the open ocean tidal streams [illegible] reds of about 10 cm/sec (less than 0.2 knot). Speeds increase wherever the cross section of the stream becomes narrower. In the southern North Sea, the English Channel, and the Gulf of Maine, speeds of 1.0 m/sec (2 knots) occur in some straits. In the Pentland Firth and between the Lofoten Islands tidal current speeds are as great as 4 5 m/sec (9 knots) During stormy weather, waves on the surface may run in a direction opposed to the tidal stream and produce much feared, short, breaking waves, for example, the Maelstrom (the Moskenstraumen near the Lofotens). These waves make navigation by small ships very dangerous

Strong tidal streams have an erosive effect They keep open, for instance, the channels in tidal marshes, the sea gates between the Frisian Islands, and the shipping lanes in tidal rivers. They also prevent the formation of shelves and deltas at the mouths of rivers and help to further the formation of estuaries Deltas are generally found in waters having few strong tidal streams (for example, in the Baltic Sea, the rivers Vistula and Nogat; in the European Med[illegible]

Mississippi, Magdalena, and Orinoco). Very strong tidal streams can change the positions of sandbanks and thereby alter the course of the water channels In order to ensure the safety of sea passages in tidal rivers, depth surveys must be made at regular intervals, especially in shipping lanes. Strong tidal streams develop very high energy. There have been many attempts to utilize this energy through tidal barrages, which can be used for the production of electricity. For a discussion of power from tides see WATER POWER. [G.DI.]

MEASUREMENT OF CURRENTS AND MIXING

Many types of instruments and devices are used for measuring ocean currents, probably more than for any other single oceanographic measurement. Methods and techniques employed vary greatly too, depending upon such factors as the use of floating or attached measuring devices, the depth at which observations are desired, and the need for detailed measurements at a given location or numerous measurements over large areas. This discussion considers some of the more commonly used methods of measuring both ocean currents and the deep circulation and mixing that take place within the oceans

Direct methods of measurement. Current-measuring devices are of several functional types. The first utilizes the drift of a free body such as a drogue, a drift bottle, a ship, or a mid-depth neutrally buoyant pinger Speed and direction are determined by observing the distance and direction the body drifts in a given time interval The second method is based on drag effects on a fixed body. Information can be obtained when the current rotates a propeller, twists a vane, tilts an instrument case, or creates a pressure difference in a pitot tube In both methods it is necessary to know the actual motion of the instrument or to know that it

Fig. 6b. The level of the sea is higher than in the neighboring ocean. The surface water with its lower density and accordingly its lower salinity therefore flows outward and the relatively salty ocean water of higher density flows over the sill into the marginal sea (examples of this circulation are the Baltic Sea with the shallow Darsser and Drogden rises, the Norwegian and Greenland fiords with fiord rises, and the Black Sea with its entrance through the Bosporus). Should the sill depth always remain in the water of low density (outflow of the upper layer) as is the case with the Bosporus, then the renewal of deep water and with it the circulation of deep-sea water in the marginal sea come to a complete halt. The result of this, with respect to the water constituents, is that the oxygen is entirely used up and poisonous hydrogen sulfide takes its place. For this reason, below depths of 200 m no life is possible in the Black Sea. If the sill depth interferes only occasionally with the lighter water of the upper levels, as in the entrances to the Baltic Sea, then the renewal of deep water is interrupted only at intervals. In the Baltic Sea these interruptions sometimes last for several years. *See* BALTIC SEA; BLACK SEA; FIORD.

Deep circulation in the oceans. The deep circulation in the oceans is more difficult to perceive than the circulation in the marginal seas. In addition to the internal pressure forces, which are conditioned by the distribution of density and the effect of the piling up of water caused by the wind, there are also the influences of Coriolis forces and turbulence which must be considered. Also there are areas in which the surface water, as a result of climatic conditions, takes on a relatively high density. In thermohaline convection the water sinks until it reaches a layer in which there is a density corresponding to its own and then spreads out horizontally. In this way the cold and deeper levels of the oceans, the so-called cold-water sphere, take on a leaflike or layer structure consisting of bottom water, deep water, and intermediate water. In the Atlantic Ocean where this distribution of water was systematically examined by A. Defant (1941) and G. Wüst (1957), the deep-water circulation is strongly marked on the western side of the ocean, where measurable speeds are found. Figure 7 is a schematic representation of the surface and deep circulation in the Atlantic Ocean as viewed from the western side of the ocean. From the lines indicating equal salinity, the origin of the water masses in the various strata can be concluded. The Bottom Water comes from very cold water masses which have sunk along the edge of Antarctica. In the layer immediately above, or Deep Water, the water masses have their origin south of Greenland. In the next layer above, or Intermediate Water, the water masses come from the polar front in the Southern Hemisphere.

There are five areas where the surface water becomes denser and sinks. It is in these areas that the circulation of the deep sea has its origin.

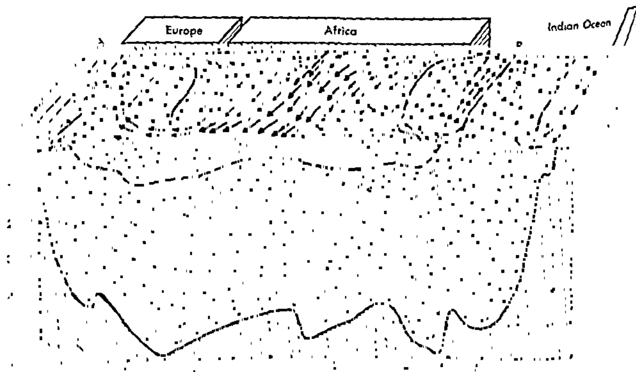


Fig. 7. Schematic representation of the surface and deep circulation in the Atlantic Ocean. All arrows show current directions; on the surface, thin arrows indicate speeds of 5–40 cm/sec (0.1–0.8 knot) and thick arrows indicate speeds of 40–150 cm/sec (0.8–2.9 knots). SC indicates convergence of surface currents in subtropical waters. P indicates oceanic polar front where cold-water masses from polar and subpolar geographical latitudes meet relatively warm waters of temperate zone. In vertical section heavy broken line shows division between warm- and cold-water spheres, and other lines indicate equal salinities. (After G. Wüst, 1949)

self contained Ekman propeller-type meters which record the revolutions of a propeller in a given length of time, as well as the direction by compass (Fig 8). There are propeller type meters which are self contained, and there are meters which telemeter their data to the ship or a surface buoy. The Roberts meter is an American version of the latter type (Fig 9). In weak currents an S-shaped Savonius rotor is less affected by vertical motion than a conventional propeller. Underwater cameras are being used increasingly to photograph deflection of current vanes, compasses, and propellers By using a slow-setting gelatin, the Carruthers meter indicates the tilt of a tube and the compass reading at time of cooling Meters of the fixed-instrument type have been useful in measuring current speeds of a few tenths of 1 knot

Indirect methods of measurement. Currents can be measured indirectly by the rate of cooling of hot wires or thermistors Another electrical method is the GEK (geomagnetic electrokinetograph) The GEK measures the potential difference between two separated electrodes which results from movement of the conducting salt water through the earth's vertical magnetic field With electrodes spaced 100 m apart a voltage of about 1 millivolt is induced per knot of current This method requires uncertain corrections in shallow water and requires that the ship make jogs in its course to measure both east-west and north-south current components

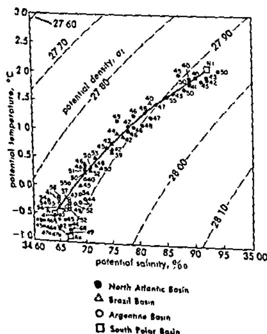


Fig 10 Relation-diagram (T-S curve) between potential temperature and salinity of the Antarctic Bottom Water in the Western Atlantic for calculating the percentage amount of the original water types A_1 = Antarctic Bottom Water, N_1 = North Atlantic Deep and Bottom Water (After G Wüst)

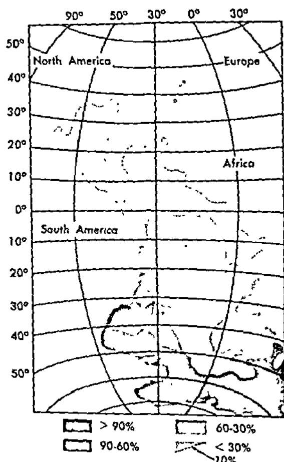


Fig 11 Spreading and mixing processes within the Antarctic Bottom Water, represented by lines of equal amount of the original Antarctic component, A_1 of Fig 10 (After G Wüst)

The subsurface flow patterns between water masses can be determined by tracing the distribution of conservative properties such as salinity along lines of constant density The quantitative calculation of geostrophic currents is made by the formulas

$$u = \frac{-1}{2\rho\omega \sin\varphi}\frac{\partial p}{\partial y} \qquad v = \frac{1}{2\rho\omega \sin\varphi}\frac{\partial p}{\partial x}$$

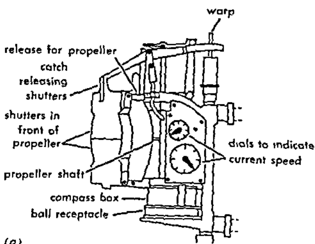
where u and v are the east and north components of velocity, x, y, and z are positive in the east, north, and downward directions, ρ is the density of sea water, p is pressure, ω the angular rotation of the earth, and φ is latitude, positive in the Northern Hemisphere In practice, temperature, salinity, and pressure are measured at various depths along a line or over a grid, density is calculated from the equation of state; and pressure is calculated from the hydrostatic equation ($dp = \rho g\, dz$), by integrating over some arbitrary depth, usually 1000 m or more (see discussion in preceding section on dynamics of ocean currents) As in meteorology this method has been of great importance although it is subject to some serious limitations

[A.C.V ; J.A K]

is stationary. Hence the navigational problem of knowing accurately the position of the current meter or ship is one of the most difficult parts of ocean-current measurement. The other great problem arises because the speed of ocean currents is usually much less than 1 knot (1.1 mph, or 51 cm/sec). Only in a few great currents such as the Gulf Stream are there speeds as great as 5 knots.

If the ship is close enough to land to be in an electromagnetic navigational network such as Loran or Decca, positions can be known to within 1 mile or even 100 yd. If such a network is not available or if the current is very slow, the position of the current meter, the ship, or a marker buoy must be maintained by anchoring. Ships and buoys can be anchored in very deep water and allowance can be made for swinging on the anchor line See NAVIGATION; NAVIGATION SYSTEMS, ELECTRONIC.

Drift of a free body. Most ocean surface currents have been discovered because of their effect on the course and speed of ships Even slow surface currents have been observed and measured by the drift of debris or drift bottles. Such observations give only average surface currents, but they are



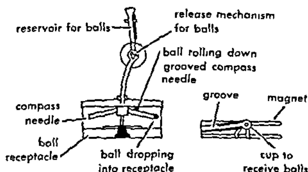


Fig. 8. Diagram of Ekman current meter (a) Meter without tail vane (b) Internal mechanism. (c) Grooved compass needle. As propeller rotates, balls fall one at a time, onto grooved compass needle, which guides them into chambers in receptacle, depending on the heading of the instrument. Dial readings give number of shaft revolutions (*Adapted from H. Barnes, Oceanography and Marine Biology, Allen & Unwin, 1959*)

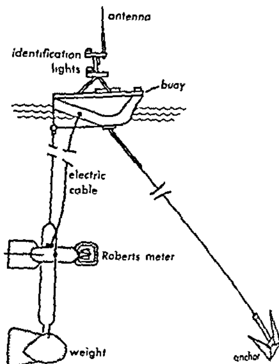


Fig. 9. Buoy-operated Roberts radio current meter Buoy contains battery-operated radio transmitter and antenna. When meter is lowered directly from ship, there is no need for the transmitter and receiver as the electric cable is connected directly to the relay box. (*Adapted from U.S Navy Hydrographic Office Publ 607, 2d ed., 1955*)

still an important part of oceanography. A drift method is advantageous in that measurements can be extended over many days, and thus the effects of tidal currents can be averaged out.

The Swallow-type neutrally buoyant float is a mid-depth current meter consisting of an aluminum pressure case which, being less compressible than sea water, can float at a predetermined depth. From a self-contained sound source it emits acoustic pings which can be heard for several miles from a quiet ship equipped with appropriate sound gear. The sensitivity of this method can approach 0.01 knot if the buoy is followed for several days It provides the most accurate way to measure mid depth currents. Development of greater range and better depth-determination gear for use with the neutrally buoyant float will make it still more useful in oceanography.

Mid-depth drift measurements are also made by a weighted aviator's parachute at depth connected by a thin wire to a small surface float. The float can be followed visually or tracked by radar This system permits good shear measurements but becomes less effective for very deep or very slow currents.

The use of marker dye or radioactive material can show how a body of water actually moves and mixes in three [illegible] potentially accu[illegible] [illegible]t to date has re[illegible]

Effects on a fixed body In this method the effect of the current on a fixed instrument is measured Most subsurface currents have been measured with

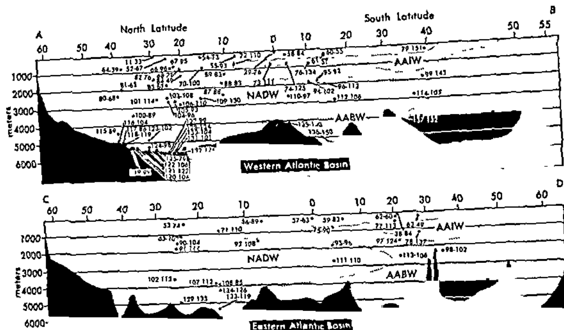


Fig 13 Profiles of radiocarbon content in Eastern and Western Atlantic Basins Number on left is sample number shown in Fig 3; number on right is the per mile difference in radioactivity with respect to 1890 wood after corrections have been made for isotopic fractionation. Note that the radiocarbon content of Antarctic Bottom Water increases from south to north (down current) by mixing with the North Atlantic Deep Water; values of 150 in the equatorial Atlantic gradually approach values of 100 near 40°N This compares favorably with the data of the *T-S* curve shown in Fig 10 and the chart in Fig. 11.

north latitude (see Figs. 12 and 13). *See* RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS, RADIOCARBON DATING, TRITIUM. [B.C H , B J.C.]

Bibliography W. S Broecker, M Ewing, R Gerard, and B C. Heezen. Geochemistry and physics of circulation, *Proc. Int Oceanographic Congress*, vol 2, 1960, G. Dietrich, *Allgemeine Meereskunde*, 1957; J. Proudman, *Dynamical Oceanography*, 1953, H Stommel, *The Gulf Stream*, 1958, H U. Sverdrup, M W. Johnson, and R H Fleming, *The Oceans*, 1942, G Wust. Die Stratosphare. *Wiss. Ergeb Deutsch Atlant Exped Meteor, 1925–1927*, 6(1):109–288, 1936

Ocean waves

The irregular moving bumps and hollows on the ocean surface Winds blowing over the ocean, in addition to producing currents, create surface water waves called waves or a "sea" (Fig 1) The characteristics of these waves (or the state of the sea) depend on the speed of the wind, the length of time and the distance over which it has blown, and on the depth of the water. If the wind dies down, the waves that remain are called a dead sea Waves can travel hundreds and thousands of miles from where they were generated into areas where the wind is light. These waves are called *swell* The rise and fall of the water as a function of time at a fixed point can be recorded Such a record, as taken by the Ocean Weather Ship *Weather Explorer* with a ship-borne wave recorder, is shown in Fig 2. The highest wave in this record is 46 ft from crest to trough. *See* SEA STATE.

This article treats the generation of sea waves, the mathematical theory of forecasting ocean waves, and the instruments used to measure ocean waves. For a discussion of wave characteristics, *see* WAVE MOTION IN LIQUIDS; *see also* SEICHE, SHORE PROCESSES; STORM SURGE, TSUNAMI, WAVE (CAPILLARY); WAVE (INTERNAL).

Generation of sea waves. The physical processes by which waves are generated are not completely understood The mathematical equations governing the motions of the irregular wavy surface have

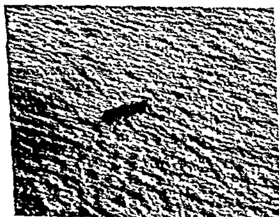


Fig 1. Aerial photograph of sea waves (*Capt D, B MacDiarmid, USCG*)

Deep circulation and mixing. The waters of the deep sea can be divided into water masses on the basis of distinctive chemical and physical properties. The geographical variation of these distinctive properties within adjacent water masses allows the oceanographer to study mixing processes in the deep sea.

Variation in properties. Temperature and salinity are termed conservative properties because they do not change except through mixing. The *oxygen, phosphate, nitrate, and other biochemical* properties are termed nonconservative because the concentrations of these substances may be altered by biological production, or the oxidation of dead biological products, and thus may change independently of mixing processes Nevertheless, both conservative and nonconservative properties may be used to study mixing. The standard method for the study of mixing is the *T-S* (temperature-salinity) curve. A water mass is determined by the fact that water from this mass always lies along an equal-density line on a graph of temperature vs. salinity. In the accompanying diagram (Fig. 10) the *T-S* curve of Antarctic Bottom Water is shown. As the Antarctic Bottom Current, A_1, moves north along the Western Atlantic Basin, it gradually mixes with the overlying North Atlantic Deep Water, N_1 (Fig. 11). Thus the temperature and salinity of the water mass change continuously Contrary to expectation, the nonconservative property, oxygen, increases down current as a result of mixing with the oxygen-rich overlying North Atlantic Deep Water, consumption by biological processes being less important than mixing in this case. Most problems of water-mass identification and mixing are much more difficult than the nearly ideal case cited above Concentrations of nonconservative properties may only be used in conjunction with *T-S* curves or plots of other conservative properties Particle concentrations have recently shown great promise in the identification of water masses and probably will be useful in the study of mixing Data suggest that fine particles sink so slowly that particle content may be considered a characteristic property of water mass.

For further treatment of the distribution of properties in sea water, mixing and stirring processes, and characteristics of ocean water masses, and of biological productivity, its measurement, geographic variation, and indicator species, *see* SEA WATER; SEA WATER FERTILITY.

Radioactive isotopes. Radiocarbon (C^{14}) and tritium (H^3) are formed in the upper atmosphere and eventually enter the surface layer of sea water. Because of its short half-life (12 years) tritium is useful as a tracer only for systems of very *fast* circulation, whereas radiocarbon's 5580-year half-life makes it suitable even for the slowest current systems. Radiocarbon is produced naturally in the atmosphere and is dissolved in the oceans as carbon dioxide. In an ideal situation, once the water leaves the surface, it has no new supply of

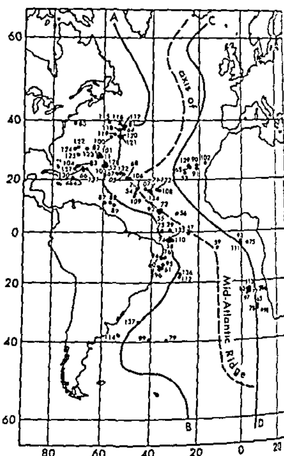


Fig. 12. Location of subsurface radiocarbon samples, Atlantic Ocean

this isotope and the dissolved C^{14} atoms continue to decay with a 5580-year half-life.

If the circulation of deep water masses were accompanied by little mixing between adjacent water masses, and the surface C^{12}/C^{14} ratio were the same everywhere, it would be possible to measure the rate of circulation in a deep current by measuring the downstream decay of C^{14} relative to C^{12} along the current axis. However, neither of these assumptions can always be made since upwelled water may sink before its C^{12}/C^{14} ratio attains equilibrium with the atmosphere and mixing between adjacent water masses is frequently very large. For example, the Antarctic Bottom Current mentioned in regard to the *T-S* curve method appears to grow younger down current because of the younger C^{12}/C^{14} ratio of the overlying water which is continuously being mixed into the water mass Thus, if the ratios are converted to ages without considering mixing, the current would appear to be moving slowly southward instead of rapidly northward

However, if the C^{12}/C^{14} values in the core of a relatively thick water mass such as the North Atlantic Deep Water are compared with the same ratio found in the surface source area of the water mass, mixing with other water masses can be neglected and an average residence time can be determined. Calculated by this method the average residence time of the North Atlantic Deep Water in the core of the water mass is 560 years at 30°

$$E[\eta(x,y,t)\eta(x+x^*,y+y^*,t+\tau)]$$
$$= \frac{1}{2}\int_0^\infty \int_{-\pi}^{\pi} S(\mu,\theta) \cos\left[\frac{\mu^2}{g}(x^*\cos\theta + y^*\sin\theta) - \mu\tau\right] d\theta\, d\mu \quad (5)$$

It was shown by W J Pierson in 1955 that the n random variables $\eta(x_1,y_1,t_1)$, $\eta(x_2,y_2,t_2)$, . . . , $\eta(x_n,y_n,t_n)$ have a multivariate normal distribution. $E[\eta(x_j,y_j,t_j)]^2$ is given by Eq. (5) with x^*, y^*, and τ equal to zero The covariances, $E[\eta(x_j,y_j,t_j)\,\eta(x_k,y_k,t_k)]$, are given by Eq. (5) with $x^* = x_j - x_k$, $y^* = y_j - y_k$, and $\tau = t_j - t_k$

By virtue of the ergodic theorem, the time and space variation of a particular sample function has the same statistical properties as the variations in the ensemble. A given sample function can be analyzed so as to estimate the spectrum and also the required parameters for the various probability density functions that describe the statistical properties of the waves

With τ equal to zero, the Fourier inversion of Eq. (5) yields information on the directional spectrum if the transformations $\alpha = \mu^2 \cos\theta/g$ and $\beta = \mu^2 \sin\theta/g$ are made, where α and β are wave numbers in a cartesian coordinate system There is a 180° indeterminancy in direction that can be resolved by an analysis of the meteorological conditions that generated the waves When a finite area is used to obtain such an estimate, the analysis of the data cannot be based simply on the above equations The techniques developed by J W. Tukey in 1949 and summarized by R B Blackman and Tukey in 1958 for single-variable cases have been extended to the two-variable case.

As shown in Fig 3, such a directional spectrum has been estimated from stereophotogrammetric measurements of waves generated by a wind of 18 7 knots near the surface Figure 3 is a smoothed version of such an analysis based on estimates that are distributed according to χ^2 (chi square probability density function) with 19 degrees of freedom. The contours when divided by 10^4 and interpreted in terms of square feet estimate the result of integrating $S(\mu,\theta)$, approximately over a square with a length of side given by the distance between two of the scale marks on the side of the figure A wide range of directions and of wavelengths from 600 ft down to 60 ft is evident The secondary peak appears to be due to swell There are even shorter waves which could not be detected by the process of analysis If the contour values are halved, Fig 3 can be interpreted as the resolution of the total variance of the wavy surface into contributions from different wave numbers

The three-variable process has many interesting properties that were studied in detail and reported upon by M S Longuet-Higgins in 1957 Such properties as the probability density function (pdf) of the speeds of contours of constant elevation, the appearance and disappearance of maxima and points of inflection, and the number of relative maxima and minima on the surface have been determined

If the waves are observed at a fixed point as a function of time, as in Fig 2, x^* and y^* become zero in Eq. (5) and directional effects are lost.

$$S(\mu) = \int_{-\pi}^{\pi} S(\mu,\theta)\, d\theta \quad (6)$$

Such a time record becomes a sample from a stationary Gaussian process, and the results of the analysis of such processes can be immediately applied to the study of wave records The estimate of the spectrum of the record shown in Fig. 2 is given in Fig 4, for example, after at least partial correction for instrumental response Each estimate has about 11 degrees of freedom Also the variation of $\eta(x,y,t)$ along any line in x, y, t space is similarly represented by a one-variable process.

The needed parameters for many theoretical pdfs that describe quantities that can be evaluated from the record of the waves can be determined from the wave spectrum. The average time interval between zeros and the average time interval between maxima in the record can be found. The crest-to-trough wave heights are roughly distributed according to a Rayleigh distribution. For pressure records made below the surface, the pdf of the zero intervals can be found fairly accurately.

Such quantities as sea surface slopes and curvatures can also be evaluated These quantities depend on the various moments of the spectrum, on the values of $E[\eta(x_1,y_1,t_1)\,\eta(x_2,y_2,t_2)]$ and on spectra derived by transformations and differentiations. See DISTRIBUTION (PROBABILITY); FOURIER SERIES AND INTEGRALS; STATISTICS

Nonlinear considerations. If the wavy surface were a truly linear process, as described by Eqs (1), (2), and (3) when linearized, these results would be expected to verify for all such operations However, the equations governing the wave motion are truly nonlinear In 1958, L. J. Tick obtained the correction to the Gaussian model that must be considered in studying second order nonlinear effects for the special case of long-crested progressive waves If p is zero in (2), and if all partials in y are set equal to zero in (1), (2), and (3), and if $\eta^{(1)}(x,t)$ and $\phi^{(1)}(x,t)$ satisfy (1), (2), and (3) when linearized, then $\eta^{(2)}(x,t)$ and $\phi^{(2)}(x,t)$ can be found such that (1), (2), and (3) are satisfied to second order for $\eta(x,t) = \eta^{(1)}(x,t) + \eta^{(2)}(x,t)$ and $\phi(x,t) = \phi^{(1)}(x,t) + \phi^{(2)}(x,t)$ If $S^{(1)}(\mu)$ is the spectrum of $\eta^{(1)}(x,t)$, then $S^{(2)}(\mu)$ is given by (7). $S(\mu) = S^{(1)}(\mu) + S^{(2)}(\mu)$ is the spectrum of a non-Gaussian random process.

$$S^{(2)}(\mu) = \frac{1}{g^2}\int_{-\infty}^{\infty} K(\lambda,\mu) S^{(1)}(\mu-\lambda) S^{(1)}(\lambda)\, d\lambda \quad (7)$$

where

$$K(\lambda,\mu) = \begin{cases} \lambda^2(\mu^2 - 2\mu\lambda + 2\lambda^2) & 0 < \lambda < \mu,\ \mu > 0 \\ (\mu - 2\lambda)^2\mu\lambda & \lambda < 0,\ \lambda > \mu,\ \mu > 0 \end{cases} \quad (8)$$

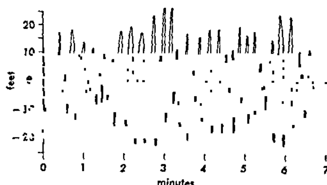


Fig. 2. Record from ship-borne wave recorder taken November 16, 1953 at 60 2°N, 14.0°W The wind was Beaufort force 11. (National Institute of Oceanography, Wormley, Surrey, England)

never been completely solved, as they are nonlinear. Energy is supplied to the waves by the tangential and normal stress components of the wind. Part of this energy is dissipated in the wave motion by eddy viscosity and by the breaking of wave crests. The large values of the eddy viscosity may be due to the effect of breaking waves. A fully developed state will exist if the energy input equals the energy dissipated. The waves will grow if the energy input exceeds the energy dissipated.

Any physical theory for calculating the growth of the waves under the action of wind must be based on a thorough knowledge of both the energy transfer from wind to waves and the rate of dissipation of wave energy in different phases of wave development. So far no theory exists which satisfactorily explains the growth of the composite wave motion under wind action. For practical purposes of wave forecasting, several methods have been suggested which are based on direct observations

Equations of motion. If potential flow can be assumed, and if the effect of surface tension is neglected, the equations governing the motions of the water are described by the potential equation (1), Bernoulli's equation (2), the kinematic boundary condition (3), and the condition that the potential $\phi_z \rightarrow 0$ as $z \rightarrow -\infty$ if the water is deep (subscripts denote partial differentiation). The equations governing the motions of the air above should also be considered, but are not given here. The position of the boundary between the water and the air is described by $z = \eta(x,y,t)$, the free surface, where z is positive upward.

$$\phi_{xx} + \phi_{yy} + \phi_{zz} = 0 \qquad (1)$$

$$\frac{p}{\rho} + gz - \phi_t + \tfrac{1}{2}(\phi_x^2 + \phi_y^2 + \phi_z^2) = 0 \qquad (2)$$

at $z = \eta(x,y,t)$ and below the surface

$$\eta_t = -\phi_z + \phi_x\eta_x + \phi_y\eta_y \qquad \text{at } z = \eta(x,y,t) \qquad (3)$$

With the wind blowing, the pressure p at $z = \eta(x,y,t)$ will be a function of x, y, and t. If p at $z = \eta$ is set equal to zero and if the squared and product terms of the above equations are omitted, the equations become linear. Numerous solutions exist. The important problem is to find solutions that approximately represent waves as found in nature.

Wave statistics and stochastic models With complicated computations and sufficiently detailed recording techniques, the wavy surface over an finite area could be represented as closely as desired by a mathematical formula Wave records of any length, such as the one in Fig 2, could be represented as closely as desired as a function of time, but such an effort would be wasted The space pattern could be observed only over an area that is small compared to the dimensions of the oceans The time record could be observed only for a short period. The waves would never again be exactly like the ones that were studied in such detail.

The waves must therefore be studied by means of statistical models, and given wave records or stereophotogrammetric observations must be analyzed by means of statistical techniques This permits economy of computation. The analysis of a given record that will never again be repeated then fits into an over-all pattern that describes the statistical properties of the waves. Fortunately, such models exist, from the field of electronics in particular, and they have been extended and adapted to the study of ocean waves.

Such a model is given by the ensemble of all possible sea states for a particular power spectrum, $S(\mu,\theta)$, as in Eq. (4).

$$\eta(x,y,t) = \int_0^\infty \int_{-\pi}^{\pi} \cos\left[\frac{\mu^2}{g}(x\cos\theta + y\sin\theta) - \mu t + \epsilon(\mu,\theta)\right]\sqrt{S(\mu,\theta)\,d\mu\,d\theta} \qquad (4)$$

Integrals of this type were defined, for example, by P. Levy in 1948 The symbol $\epsilon(\mu,\theta)$ stands for a random phase uniformly distributed over the interval between zero and 2π. The power (or energy) spectrum, $S(\mu,\theta)$, has the dimensions of cm² sec/radian If Eq (4) is represented by an approximating double summation, the sea surface can be represented by a large sum of many simple harmonic progressive waves, each with a different frequency $\mu = 2\pi/T$, a wavelength determined by $2\pi/\lambda = \mu^2/g$ $(\lambda = gT^2/2\pi)$, a phase speed equal to $c = g/\mu$ $(c = gT/2\pi)$, and a direction toward which the wave is traveling determined by θ. The amplitude of the wave is determined by the square root of the volume under $S(\mu,\theta)$ for the appropriate range of μ and θ. Other representations are also possible

An ensemble of sea states can be defined by considering ${}^{(1)}\eta(x,y,t)$, ${}^{(2)}\eta(x,y,t)$, . . . , ${}^{(n)}\eta(x,y,t)$, ${}^{(n+1)}\eta(x,y,t)$, . . . , where the only difference from record to record would be the random phases chosen for each term in each partial sum

For the ensemble space, the covariance function can be found as in Eq (5), where E denotes the expected value

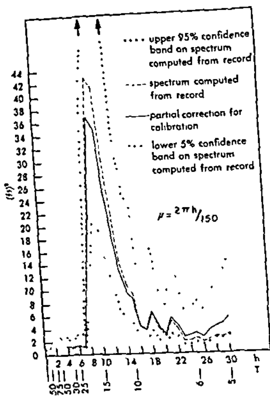


Fig 4 Frequency spectrum of the wave record shown in Fig 2

These equations indicate that results obtained from the linear Gaussian model should be questioned when they involve fourth or higher moments of the spectrum and slopes and curvatures of the sea surface The second-order nonlinear effects propagate by means of a convolution on the "assumed" spectrum of the linear part to high frequencies in such a way that these features cannot be interpreted correctly in a linear model The study of slopes and curvatures emphasizes the high frequencies where surface tension, which has been neglected in the above derivation, and nonlinear effects become more important

Wave forecasting. Sea state forecasts are based on a combination of empirical knowledge and certain theoretical relationships of waves and wind Since the sea surface wave pattern usually consists of a locally generated sea and superimposed swell which was generated in a distant storm, two separate forecasting procedures are necessary, both generally based on oceanic weather maps After combination of these two independent phenomena, the actual state of the sea surface pattern at a given locality can be estimated The most rational way of describing the state of a sea is provided by the wave energy spectrum, $S(\mu,\theta)$

The first part of the problem of wave forecasting is to find the energy spectrum as it changes from hour to hour or from place to place as the wind blows over the ocean surface. The growth of the spectrum of the sea in an area of wave generation has been described by various authors, whose hypotheses and results do not always agree with each ... the higher the ... eater ... tch), the higher the waves; and that the longer the wind blows (the duration), the higher the waves. The upper limit of wave height for a given wind is not agreed upon, nor are the fetches and durations that produce maximum development.

The second part of the problem of forecasting waves is the problem of describing how the wave height decreases when the wind dies down, and how the waves travel out of the area of generation as swell Numerical filtering operations (somewhat analogous to electronic filters) that depend on the dimensions of the generating area and on the time of the forecast can be applied to the spectrum of ...

that the waves are highly dispersive. The spectrum $S(\mu,\theta)$ covers a wide range of frequencies and directions, and therefore the waves spread out over a wide area as they travel out of the generating area

The effect of viscosity does not seem to be important unless the waves have to travel through another storm area, where the eddy viscosity in the water is significant *See* HYDRODYNAMICS, SURFACE TENSION [C.N.; W.J.P.]

Wave-measuring devices. Instruments designed to measure ocean waves can be grouped into two classifications: those that sense the elevation of the surface water, and those that sense the subsurface pressure fluctuations generated by the waves. Surface elevation sensing gages include (1) surface float devices mechanically connected to a recording mechanism, (2) devices that record the buoyant force on a vertical cylinder, (3) recorders actuated by vertical accelerometers mounted in surface buoys, (4) inverted echo sounders (fathometers) fixed below the surface to echo off the water surface, (5) accurate absolute altimeter recorders operated in aircraft flying at a fixed elevation, (6) stereophotographs, and (7) electrical elements whose resistance or capacitance is a function of the elevation of the water surface.

Surface gages Of the gages that sense the surface elevation, the step resistance gage shown in Fig. 5 (type 7 above) is the most widely used. Electrical contact points are mounted along a vertical support and connect to a resistance circuit. The values of the resistors connected to the contact points are selected so that the current increases in proportion to the number of contacts shorted along the submerged length of the gage. The alternating current (used to prevent polarization of the contacts) which flows through the gage is converted by means of a rectifier to a proportional direct current to drive a pen recorder.

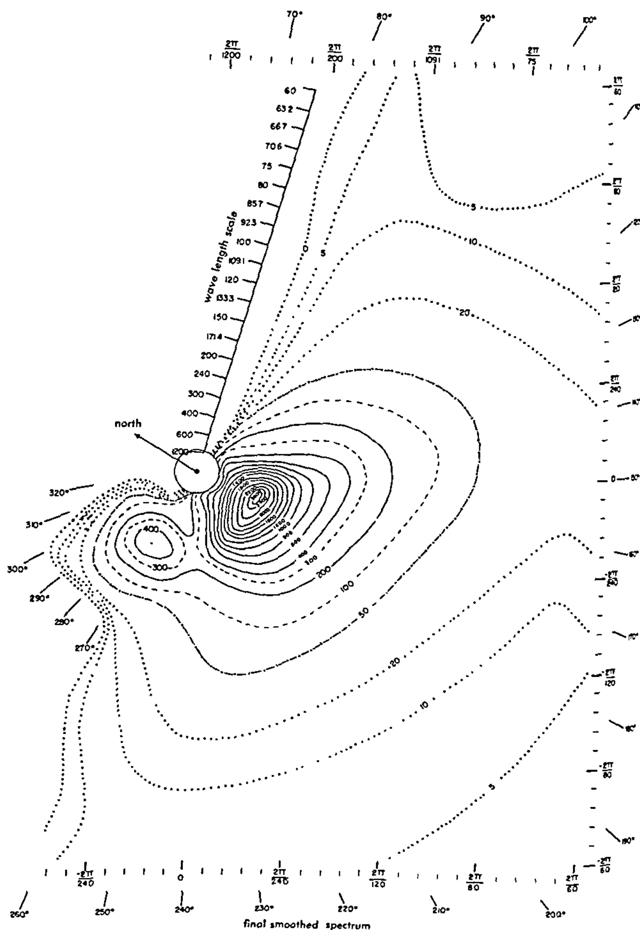


Fig 3. Directional spectrum of a wind-generated sea. *(From NYU Meteorol Papers)*

Oceanic islands

Those islands which rise from the deep-sea floor rather than from shallow continental shelves Most islands in gulfs and seas that fringe the great ocean basins are geologically similar to the nearby continents On the other hand, almost all islands that rise from the ocean basins are volcanoes with or without coral reef and, geologically, bear little relation to the continents Volcanic islands are only the tops of much larger undersea volcanoes, most of which are associated with great submarine structures, such as submarine ridges and fractures in the earth's crust (Fig 1)

Submarine volcanoes. On the deep-sea floor volcanoes begin as lava flows from fissures in the earth's crust under two or three miles of water. Gradually they build upward through the water, but about nine-tenths of the submarine volcanoes become inactive and stop their growth before they reach the sea surface. The others burst from the deep water into a new realm where wave and subaerial erosion combat their upward growth. At first the volcanoes tend to produce ash and cinders which are easily eroded. Falcon Island, an active volcano in the Tonga group, has several times been reduced to a submarine bank within a few years after an eruption built up an island. Gradually as the pile becomes broader, volcanoes rise above the waves and more resistant fluid lava flows build a solid island. Where several nearby volcanoes merge together, as in Hawaii, a great island may form. See VOLCANO; VOLCANOLOGY.

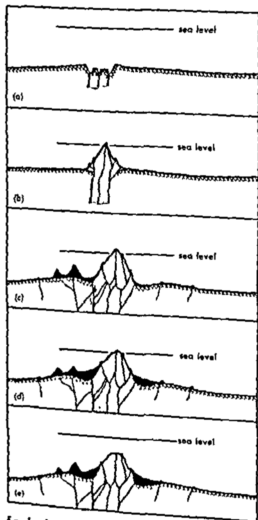


Fig 1 Birth and development of submarine volcano or ridge Gray areas represent initial extrusion of volcanic material; black areas indicate subsequent deposits of volcanic material and erosional debris Sequence (a) to (e) explained in text

Volcanoes are active for no more than a few million years, however, and inactive volcanoes are inevitably worn down to shallow submarine banks by erosion which never stops. In addition to these worn down volcanoes, drowned former islands called guyots or tablemounts have been discovered in all the ocean basins, mostly at depths of 1000–7000 ft. Reef coral as old as the Cretaceous and volcanic erosional debris have been dredged on some guyots. Also, drilling on atolls shows that the coral is a capping several thousand feet thick on former volcanic islands See ATOLL; SEAMOUNT AND GUYOT.

Associated submarine structures. Most submarine volcanoes are associated with great submarine structures: long, straight, narrow features such as the Hawaiian Ridge and Murray Fracture Zone; broad oceanic rises such as the Mid-Atlantic Ridge; and island arcs and trenches such as the Aleutian Arc (see SUBMARINE TOPOGRAPHY).

Long straight structures Lines of volcanoes occur in the Atlantic and Indian Oceans but are relatively rare A linear group of guyots extends southeast from Cape Cod, and other groups may be undiscovered in the less-well-surveyed parts of these oceans.

It is the Pacific, however, that is the type area for linear archipelagoes, and there they are extremely common. Existing linear groups are largely confined to the southwestern and central Pacific (Fig. 2), but, in the past, islands were present in the northern part of the basin (Fig 3). Some very large archipelagoes like the Tuamotu Islands and former archipelagoes such as the Mid-Pacific Mountains consist of individual volcanoes only a few thousand feet high rising as peaks above great steep-sided ridges Other archipelagoes including the Hawaiian, Samoan, and Marquesas Islands have large volcanoes rising from lower ridges ...

... [illegible] along the Clarion Fracture Zone, a submarine feature in the eastern central Pacific A long straight trough forms the western part of the fracture zone. Toward the east it is interrupted by seamounts until the trough disappears and is replaced by a line of shallow banks,

Wave-measuring instruments that sense subsurface pressure fluctuations rely on hydrodynamic theory to compute the surface wave heights. The equations are as follows:

$$K = \frac{\cosh \dfrac{2\pi b}{L}}{\cosh \dfrac{2\pi d}{L}} \quad \text{and} \quad L = \frac{g}{2\pi} T^2 \tanh \frac{2\pi d}{L} \qquad (9)$$

where K is ratio of pressure variation expressed as equivalent water heights to surface wave height, d is the depth of the water, b is the height of the instrument above bottom, L is the wave length, T is the wave period, and g is the acceleration of gravity. For example, if $d = 40$ ft, $b = 10$ ft and the pressure record indicates $T = 10$ sec with an amplitude of 5 ft, $g/2\pi = 5.12$ ft/sec², L would be 329 ft, K would be 0.817, and the surface wave would have an amplitude of 5 0/0 817 = 6.1 ft

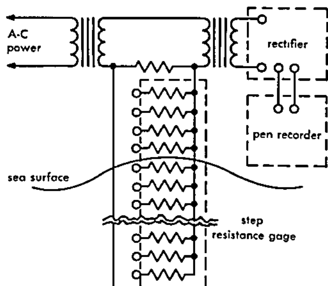


Fig 5. Step resistance wave-measuring gage

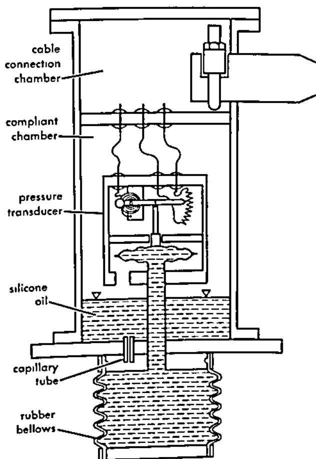


Fig. 6 Subsurface pressure gage.

Subsurface gages. Subsurface pressure signals can be converted into an electrical signal by any of the numerous transducers A typical pressure gage which utilizes a differential potentiometer type transducer is shown in Fig 6. One pressure port of the transducer is connected to the sea pressure by means of silicone oil and a rubber bellows. The second port is open to the air in the compliant chamber.

As a result of the restricted flow of fluid into this chamber through the capillary tube, the pressure is equal to the average pressure The differential pressure across the transducer is therefore only that produced by the wave-generated pressure fluctuations. The mean pressure (hydrostatic) and the slow pressure fluctuations due to tides are cancelled and do not appear in the record [F.E.S.]

Bibliography: R. B. Blackman and J. W Tukey The measurement of power spectra from the point of view of communications engineering, *Bell Syst Tech J.*, vols 1 and 2, 1958; J. Chase et al., The directional spectrum of a wind generated sea as determined from data obtained by the Stereo Wave Observation Project, *Meteorol. Papers N Y. Univ.* 2(6), 1959; J. Darbyshire, A further investigation of wind generated waves, *Deut. Hydrograph Z.* 12(1), 1959; S. Fluegge (ed.), *Handbuch der Physik*, vol. 48, 1957; M. S Longuet-Higgins, The statistical analysis of a random moving surface, *Phil. Trans. Roy Soc. London*, ser. A, 249(966): 321–387, 1957; J W. Miles, On the generation of surface waves by sheer flow, *J. Fluid Mech.*, 3(2): 185 1957; O. M. Phillips, On the generation of waves by turbulent wind, *J. Fluid Mech.*, 3(2):417 1957, O. M. Phillips, The equilibrium range in the spectrum of wind generated waves, *J. Fluid Mech.*, 4(4) 426–434, 1958; W J. Pierson, Jr., Wind generated gravity waves, in H. E. Landsberg (ed), *Advances in Geophysics*, vol. 2, 1955; W. J. Pierson, G Neumann, and R. W. James, *Practical Methods for Observing and Forecasting Ocean Waves by Means of Wave Spectra and Statistics*, U S Navy H.O Publ. 603, 1955; J. J. Stoker, *Water Waves*, 1957; L. J Tick, *A Non linear Random Model of Gravity Waves*, J. Math and Mech., 8(5):643-652, 1959; M. J. Tucker, A shipborne wave recorder, *Trans. Soc. Naval Arch. Marine Engrs London*, 98:236–250, 1956; M. J. Tucker, The accuracy of wave measurements made with vertical accelerometers, *Deep-Sea Research*, 5(3):185-192, 1959; R. L. Wiegel (ed), *Coastal Engineering Instruments*, 1955.

Oceanic islands

Those islands which rise from the deep-sea floor rather than from shallow continental shelves. Most islands in gulfs and seas that fringe the great ocean basins are geologically similar to the nearby continents. On the other hand, almost all islands that rise from the ocean basins are volcanoes with or without coral reef and, geologically, bear little relation to the continents. Volcanic islands are only the tops of much larger undersea volcanoes, most of which are associated with great submarine structures, such as submarine ridges and fractures in the earth's crust (Fig. 1)

Submarine volcanoes. On the deep-sea floor volcanoes begin as lava flows from fissures in the earth's crust under two or three miles of water. Gradually they build upward through the water, but about nine-tenths of the submarine volcanoes become inactive and stop their growth before they reach the sea surface. The others burst from the deep water into a new realm where wave and subaerial erosion combat their upward growth. At first the volcanoes tend to produce ash and cinders which are easily eroded. Falcon Island, an active volcano in the Tonga group, has several times been reduced to a submarine bank within a few years after an eruption built up an island. Gradually as the pile becomes broader, volcanoes rise above the waves and more resistant fluid lava flows build a solid island. Where several nearby volcanoes merge together, as in Hawaii, a great island may form. See VOLCANO; VOLCANOLOGY.

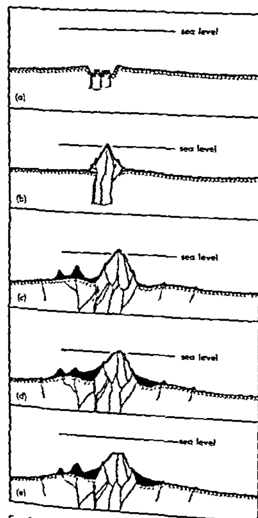


Fig. 1. Birth and development of submarine volcano or ridge. Gray areas represent initial extrusion of volcanic material, black areas indicate subsequent deposits of volcanic material and erosional debris. Sequence (a) to (e) explained in text.

Volcanoes are active for no more than a few million years, however, and inactive volcanoes are inevitably worn down to shallow submarine banks by erosion which never stops. In addition to these worn down volcanoes, drowned former islands called guyots or tablemounts have been discovered in all the ocean basins, mostly at depths of 1000–7000 ft. Reef coral as old as the Cretaceous and volcanic erosional debris have been dredged on some guyots. Also, drilling on atolls shows that the coral is a capping several thousand feet thick on former volcanic islands. See ATOLL; SEAMOUNT AND GUYOT.

Associated submarine structures. Most submarine volcanoes are associated with great submarine structures: long, straight, narrow features such as the Hawaiian Ridge and Murray Fracture Zone; broad oceanic rises such as the Mid-Atlantic Ridge; and island arcs and trenches such as the Aleutian Arc (see SUBMARINE TOPOGRAPHY).

Long straight structures. Lines of volcanoes occur in the Atlantic and Indian Oceans but are relatively rare. A linear group of guyots extends southeast from Cape Cod, and other groups may be undiscovered in the less-well-surveyed parts of these oceans.

It is the Pacific, however, that is the type area for linear archipelagoes, and there they are extremely common. Existing linear groups are largely confined to the southwestern and central Pacific (Fig. 2), but, in the past, islands were present in the northern part of the basin (Fig. 3). Some very large archipelagoes like the Tuamotu Islands and former archipelagoes such as the Mid-Pacific Mountains consist of individual volcanoes only a few thousand feet high rising as peaks above great steep-sided ridges. Other archipelagoes including the Hawaiian, Samoan, and Marquesas Islands have large volcanoes rising from lower ridges. The occurrence of volcanoes in a straight line suggests an underlying linear fracture in the earth's [illegible]

[illegible] feature in the eastern central Pacific. A long straight trough forms the western part of the fracture zone. Toward the east it is interrupted by seamounts until the trough disappears and is replaced by a line of shallow banks,

Wave-measuring instruments that sense subsurface pressure fluctuations rely on hydrodynamic theory to compute the surface wave heights. The equations are as follows:

$$K = \frac{\cosh \frac{2\pi b}{L}}{\cosh \frac{2\pi d}{L}} \quad \text{and} \quad L = \frac{g}{2\pi} T^2 \tanh \frac{2\pi d}{L} \qquad (9)$$

where K is ratio of pressure variation expressed as equivalent water heights to surface wave height, d is the depth of the water, b is the height of the instrument above bottom, L is the wave length, T is the wave period, and g is the acceleration of gravity. For example, if $d = 40$ ft, $b = 10$ ft and the pressure record indicates $T = 10$ sec with an amplitude of 5 ft, $g/2\pi = 5.12$ ft/sec², L would be 3?? ft, K would be 0.817, and the surface wave would have an amplitude of 5.0/0.817 = 6.1 ft.

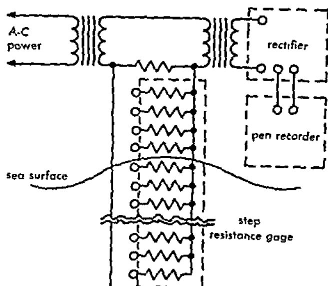


Fig. 5 Step resistance wave-measuring gage

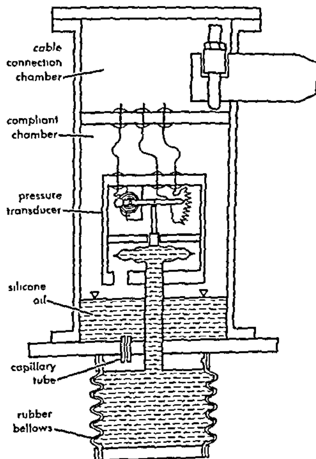


Fig. 6. Subsurface pressure gage.

Subsurface gages. Subsurface pressure signal can be converted into an electrical signal by any of the numerous transducers. A typical pressure gage which utilizes a differential potentiometer type transducer is shown in Fig. 6. One pressure port of the transducer is connected to the sea pressure by means of silicone oil and a rubber bellows. The second port is open to the air in the compliant chamber.

As a result of the restricted flow of fluid into this chamber through the capillary tube, the pressure is equal to the average pressure. The differential pressure across the transducer is therefore only that produced by the wave-generated pressure fluctuations. The mean pressure (hydrostatic) and the slow pressure fluctuations due to tides are cancelled and do not appear in the record. [F.E.S.

Bibliography · R. B. Blackman and J. W. Tukey, The measurement of power spectra from the point of view of communications engineering, *Bell. Sys. Tech. J.*, vols. 1 and 2, 1958, J. Chase et al., The directional spectrum of a wind generated sea as determined from data obtained by the Stereo Wave Observation Project, *Meteorol. Papers N.Y. Univ.* 2(6), 1959; J. Darbyshire, A further investigation of wind generated waves, *Deut. Hydrograph. Z.* 12(1), 1959, S. Fluegge (ed.), *Handbuch der Physik*, vol 48, 1957, M. S. Longuet-Higgins, The statistical analysis of a random moving surface, *Phil. Trans. Roy. Soc. London*, ser. A, 249(966): 321–387, 1957; J. W. Miles, On the generation of surface waves by sheer flow, *J. Fluid Mech.*, 3(2): 185, 1957, O. M. Phillips, On the generation of waves by turbulent wind, *J. Fluid Mech.*, 3(2):417, 1957, O. M. Phillips, The equilibrium range in the spectrum of wind generated waves, *J. Fluid Mech.* 4(4):426–434, 1958, W. J. Pierson, Jr., Wind generated gravity waves, in H. E. Landsberg (ed.), *Advances in Geophysics*, vol 2, 1955; W. J. Pierson [illegible]

1957; L. J. Tick, *A Non-linear Random Model of Gravity Waves*, J. Math. and Mech., 8(5):643–652, 1959; M. J. Tucker, A shipborne wave recorder, *Trans. Soc. Naval Arch. Marine Engrs London*, 98:236–250, 1956; M. J. Tucker, The accuracy of wave measurements made with vertical accelerometers, *Deep-Sea Research*, [illegible]:185–192, 1959; R. L. Wiegel (ed.), *C*[illegible] *ing Instruments*, 19[illegible]

and even farther to the east it is replaced by the volcanic Revillagigedo Islands.

The earth's crust is, at first, able to support the load of a new volcano or volcanic ridge, but as the structure becomes larger a point is reached when crustal strength is insufficient and elastic downbowing begins (Fig. 1b). The topographic expression is that of a central ridge surrounded by a depression or moat outside of which an arch may occur. A single volcano may be encircled by a small individual moat. As downbending continues, tension fractures permit volcanism on the arches (Fig. 1c). Volcanism and erosion debris commence filling the marginal depressions and faulting may occur. These depressions may become filled with volcanic material and sediment to form smooth archipelagic aprons sloping away from the island, seamount, or ridge, as for example around the Marquesas Islands (Fig. 1d). If eroded to a flat bank and relatively sunk below the surface, the seamount becomes a guyot (Fig. 1e), if coral was present at the surface and kept pace with the sinking, an atoll forms.

Broad oceanic rises. Submarine ridges or rises are the locus of solitary volcanic islands and seamounts such as Ascension, Reunion, and Easter Islands. Typically volcanoes are located near, but not on, the crests of the broad submarine rises. In addition, volcanic islands occur in clusters. Examples are the Azores and Galapagos Islands. The latter group is surrounded by a thin archipelagic apron, but moats have not been found around this type of island cluster.

Island arcs. Groups of islands which follow more or less curved lines are called island arcs. They are associated with deep trenches, large gravity anomalies, and earthquakes, hence they are in a region of great crustal instability. Typically there is an inner arc of active volcanoes and an outer arc of nonvolcanic islands which may contain sediments of the types now found on the deep-sea floor, thus indicating uplift of several miles. Uplift is also shown by raised sea cliffs and deposits of coral. Drowned former islands do not occur along island arcs. See TECTONIC PATTERNS. (E.L.H., H.W.M.)

Bibliography: R. S. Dietz, Marine geology of Northwestern Pacific, *Bull. Geol. Soc. Am.*, 65: 1199, 1954; E. L. Hamilton, Sunken islands of the mid Pacific mountains, *Geol. Soc. Am. Mem.* 64, 1956; P. H. Kuenen, *Marine Geology*, 1950; H. W. Menard, Archipelagic aprons, *Bull. Am. Assoc. Petrol. Geologists*, 40:2195, 1956; F. P. Shepard, *Submarine Geology*, 1948.

Ocean-meteorological relations

A focus of investigation upon the boundary zone between sea and air and upon the dynamic relationships between oceanographic studies and those of meteorology. When it is considered that nearly one half of the heat energy that the atmosphere receives for maintaining its circulation is derived from the condensation of water vapor originating primarily from oceanic evaporation, it becomes evident that an understanding of processes occurring at the air-sea boundary is fundamental to an understanding of atmospheric behavior. It will be shown that the oceanic energy supply to the atmosphere is highly regionalized owing to the character of ocean currents, which in turn implies that the atmospheric circulation itself (and resulting weather) is greatly influenced by the oceanic circulation. In the converse sense it is to be considered that the oceanic circulation represents a state of equilibrium in which the effects of the frictional stresses of the wind on the sea surface are balanced by changes in the distribution of density of oceanic waters. These compensating density changes are in turn related to time and space variations in radiation, heat conduction, evaporation, and precipitation. It is therefore equally manifest that an understanding of the oceanic circulation (and the resulting distribution of properties within the ocean) requires a thorough knowledge of atmospheric processes at the air-sea boundary.

The conclusion is that neither ocean nor atmosphere should be treated independently, but rather they should be considered together as a single dynamical-thermodynamical system. However, the interaction of ocean and atmosphere is so complicated that it is not yet possible to completely separate cause and effect. It is for this reason that separate discussions are presented for each of the major classes of atmospheric influences upon the ocean as well as the oceanic influences upon the atmosphere.

Effects of wind on ocean surface. The frictional stresses of the wind on the surface of the sea produce (1) ocean waves (and storm surges) and (2) ocean currents. The former are transitory phenomena and will not be discussed in the present article because they are of little direct meteorological interest, even though waves at sea, coastal breakers, and storm tides are of considerable maritime as well as oceanographic importance (see SEA STATE; STORM SURGE). Wind-induced ocean currents, on the other hand, are of large-scale significance from both the oceanographic and meteorological points of view.

The wind exerts a twofold effect upon the surface layers of the ocean. In the first instance, the stress of the wind leads to the formation of a shallow surface wind drift. The resulting transport of surface water by the wind drift leads in the second instance to pressure variations with depth and a changed distribution of mass (density) throughout the ocean. In the final analysis it is the resulting fields of density which account for the major current systems of the oceans. The total transport due to the wind drift is directed to the right of the wind (in the Northern Hemisphere) but the final density (slope) current which results from the sloping sea surface tends to flow in the direction of the prevailing wind except where coastal [illegible] prevents the [illegible] it should [illegible] is not the [illegible] factor that serves to

Fig. 2 Distribution of volcanic islands, banks, and atolls in the Pacific Basin.

Fig 3 Distribution of guyots, or former islands, in the Pacific Basin

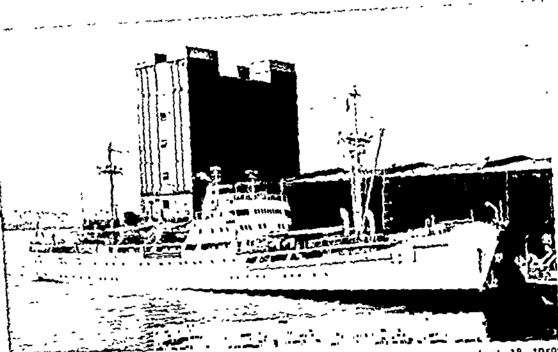


Fig 1 Soviet oceanographic ... ...ew York City. (From Electronics, Sept. 18, 1959)

ics, geochemistry, fluid mechanics, and, in its more theoretical aspects, applied mathematics Oceanography is also an environmental science which describes and attempts to explain all processes in the ocean and the interrelation of the ocean with the solid and gaseous phases of the Earth and with the Universe *See* EARTH SCIENCES

Because of the fluid nature of its contents, which permits vertical and horizontal motion and mixing, and because all the waters of the world ocean are in communication, it is necessary to study the oceans as a unit Further unification results from the technological necessity of studying the ocean from ships Many phases of oceanic research can be carried out in a laboratory, but in order to study and understand the ocean as a whole scientists ...

... scientists must go down to the greatest depths to observe and experiment (see BATHYSCAPH). Another unifying influence is the fact that many oceanic problems are so complex that their geological, biological, and physical aspects must be studied by a team of scientists Because of the unity of processes operating in the ocean and because some writers have separated marine biology from oceanography (implying the term oceanography to embrace primarily physical oceanography, bottom relief, and sediments), the term "oceanology" is sometimes used as embracing all the science divisions of the marine hydrosphere (see HYDROSPHERE) As used here the term oceanography applies to the whole of sea science.

Development of modern oceanography. The early ocean voyages by Frobisher, Davis, Hudson, ...affin, Bering, Cook, Ross, Parry, Franklin, Amundsen, and Nordenskiold were undertaken primarily for geographical exploration and in search of new navigable routes Information gathered about the ocean, its currents, sea ice, and other physical and biological phenomena was more or less incidental Later, the polar expeditions of the Scoresbys, Parry, Markham, Greeley, Nansen, Peary, Scott, and Shackleton were also voyages of geographical discovery, although scientific observations about the sea and its inhabitants were made by some of them William Scoresby took soundings and observed that discolored water containing living organisms (now known to be diatoms) was related to whale movements. Ross made dredge hauls of bottom living animals Nansen contributed to the improvement of plankton nets and suggested the existence of ...

ra ...

... Ehrenberg, whom con... of plankton ... the sea and its role in the formation of bottom deposits Charles Darwin's observations on coral reefs and Müller's invention of the plankton net belong to this phase of developing interest in marine science, in which men began to investigate ocean phenomena as biologists, chemists, and physicists rather than as oceanographers. In this group should also be such physicists and mathematicians as Kepler, Vossius, Fournier, Varenius, and Laplace, who provided the background for the development of modern theories and investigations of ocean currents and air circulation.

Toward the middle of the nineteenth century a few scientists began to study the oceans as a whole, rather than as an incidental part of an established

determine the distribution of mass or the slope of isobaric surfaces in the oceans; heating and cooling, freezing and thawing, evaporation and precipitation all exert their influences.

Major ocean current systems. For the reasons just outlined, the major ocean currents of the world conform closely with the prevailing anticyclonic wind circulations of the oceans. With the exception of the northern Indian Ocean, warm currents flow poleward to about 40° lat. in the western portion of all oceans, with easterly flow in the higher latitudes, equatorward drift (relatively cold) in the eastern portions of the oceans, and westerly flowing currents near the Equator. The low temperature of the waters in the eastern portions of the oceans is due partly to the high latitude origin of the currents and partly to coastal upwelling of cold subsurface waters. Of all the currents, the poleward-flowing warm currents of the western portions of the oceans are the best developed and the most important Examples are the Gulf Stream of the North Atlantic and the Kuroshio of the North Pacific, each of which transports a tremendous volume of warm tropical water into higher latitudes. *See* OCEAN CURRENTS.

Hydrologic and energy relations. For the earth as a whole and for the entire year, the amounts of energy received and lost through radiation are in balance for all practical purposes. However, this is not true for any given portion of the earth's surface, particularly during any given fraction of the annual solar cycle. The ratio of insolation to outgoing radiation decreases from the Equator toward the poles. Between the Equator and the 35th parallel, the earth receives more energy through radiation than it loses; the reverse is true poleward from about the 35th parallel.

Because observations indicate that the lower latitudes are not becoming progressively warmer and the higher latitudes colder, it must be assumed that considerable heat is transported from lower to higher latitudes by both atmosphere and ocean. According to H. U. Sverdrup (1953) the meridional transport of energy in the Northern Hemisphere reaches a maximum a little north of latitude 35°N. At latitude 30°N, Sverdrup computes the total energy transport across the latitude circle to be 6.5×10^{16} cal/min of which 1.9×10^{16} cal/min (or 29%) is accomplished by ocean currents, principally by the Gulf Stream and Kuroshio. The remaining energy transport is accomplished by the atmosphere.

The largest fraction of radiant energy absorbed by the oceans is utilized in evaporating sea water (*see* EVAPORATION). A much smaller fraction, about 10%, is utilized in more direct heating of the atmosphere in contact with the sea surface. The latent energy of vaporization subsequently becomes available to the atmosphere as either sensible heat or gravitational potential energy when condensation takes place, often in a region far removed from the area where the evaporation occurred The precipitation resulting from the condensation of atmospheric water vapor is then returned to the ocean, either directly as rainfall, or snowfall, or indirectly as runoff and discharge from land areas. The hydrologic and energy cycle is thereby completed. *See* MARINE INFLUENCE ON WEATHER AND CLIMATE.

The maximum evaporation and heat exchange between sea and atmosphere take place where cold air flows over warm water surfaces. The ideal locations for maximum moisture or energy transfer are therefore those areas where cold continental air flows out over warm poleward-moving ocean currents. Such ideal conditions exist during winter off the eastern coasts of the continents and over warm currents such as the Gulf Stream and Kuroshio. Radiant energy that was absorbed and stored by the oceans at lower latitudes is given off to the atmosphere by this process at places and during seasons of marked deficiency in radiative energy. Any change in the oceanic transport by ocean currents must be reflected in corresponding changes in the rates of evaporation and must finally have significant effects upon the atmospheric circulation

Relations with sea-water salinity. In the absence of horizontal flow, the surface salinity of any portion of the ocean is mainly determined by three processes: decrease of salinity by precipitation, increase of salinity by evaporation, and change of salinity by vertical mixing Salinities thus tend to be high in regions where evaporation exceeds precipitation and low where precipitation exceeds evaporation. However, horizontal transport of surface waters by wind-induced ocean currents serves to displace the areas of maximum or minimum salinity in the direction of surface flow away from the areas of maximum differences (positive or negative) between evaporation and precipitation. The conclusion, of course, is that the distributions of surface salinities (as well as other properties) of the ocean are determined almost completely by atmospheric circumstances. *See* SEA WATER.

[W. C. J.]

Bibliography: W. C Jacobs, The energy exchange between sea and atmosphere and some of its consequences, *Bull. Scripps Inst. Oceanogr. Univ. Calif.*, 6(2): 27–122, 1951. G P. Kuiper (ed.), *The Solar System*, vol. 2, 1954; H. U. Sverdrup, *Oceanography for Meteorologists*, 1942

Oceanography

The scientific study and exploration of the oceans and seas in all their aspects, including the sediments and rocks beneath the seas; the interaction of sea and atmosphere; the body of sea water in motion and subject to internal and external forces; the living content of the seas and the behavior of organisms in the sea; the chemical composition of the water; the physics of the sea and sea floor; and the origin of ocean basins and ancient seas. Hence oceanography, sometimes called the science of the seas, consists of the marine aspects of several disciplines and branches of science: geology, meteorology, biology, c[illegible] physics,

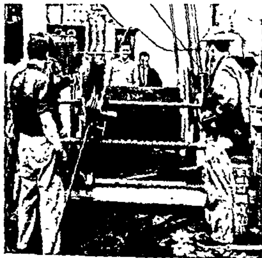


Fig 2 Trawl winch on RV Vema. (Lamont Geological Observatory)

depths. Echo sounders provide accurate profiles of the ocean floor.

Specialized equipment for oceanic exploration includes seismographs for measuring the earth's crustal thickness, magnetometers for measuring terrestrial magnetism, gravimeters for measuring variations in the force of gravity, hydrophotometers for measuring the distribution of light in the sea, heat probes (earth thermometers) for measuring the flow of heat from the earth's interior, deep-sea cameras to photograph the sea bottom, bioluminescence counters for measuring the amount of luminescent light emitted by organisms, salinometers for measuring directly the salinity of sea water, and current meters to clock the speed of ocean currents

Positioning of a ship is very important for accurate plotting of data and detailed charting of the oceans Celestial navigation is in wide use now as in the past More recent navigational aides such as electronic positioning equipment (loran, shoran, and radar) are increasing the accuracy of positioning to within several tens of yards of the ship's true position Navigational and radio communications equipment normally is situated near the captain's bridge, but often is duplicated in the scientific laboratories in order that complete communication between the ship's operators, scientists and other participating ships can be carried on at all times

Ship's laboratories Laboratories must be adaptable for a large number of operations In general, they are of two categories, namely wet and special laboratories Wet laboratories are provided with an open drained deck so that surplus sample water can be drained out on deck Such a laboratory is located near the winches used for running out and retrieving a long string of water-sample bottles (hydrocasts). Adjoining the wet laboratory are special laboratories equipped with benches for measuring chemical properties of the recovered water, and for examination of biological and geological samples. Electronics laboratories are either part of or adjacent to the special laboratory, depending upon the size of the ship Here, numerous recording devices, amplifiers, and computers are set up for a variety of purposes, such as measurements of underwater sound, measurement of the earth's magnetic and gravity fields, and seismic measurements of the earth's crustal thickness (Fig. 3).

New devices. The most promising recent development is the inertial guidance system This allows the ship's scientists to place the ship's position to within a very small degree of error and with a reliability never before realized. The development of nuclear power plants permits extended voyages without the necessity of refueling These two devices, combined in the submarines *Nautilus* and *Skate*, have made possible the first extended journey under the Arctic ice pack. Uncharted regions of the oceans are now within the reach of exploration.

Direct visual observations of the ocean depths are fast becoming a reality, both by "manned" submersibles (bathyscaph) and by television cameras. The deep-sea cameras have been developed to the point whereby motion pictures of even the deepest parts of the sea bottom can be taken (Fig 4). However, such observations yield no information about the subsurface material. Major crustal features are determined by seismic measurement The

[illegible]

... to the physical and chemical makeup of the underlying material, however, is dependent upon penetration and actual recovery. Commonly used coring devices rarely penetrate more than 10 to 20 meters (on occasion to about 33 m) below the surface. A new "incremental" coring device

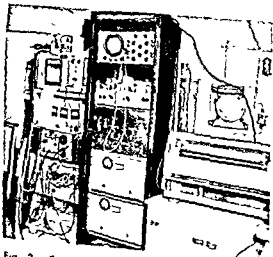


Fig 3 Special laboratory aboard research vessel (Lamont Geological Observatory)

discipline. Forbes, as a result of his work at sea, first developed a scheme for vertical and horizontal distribution of life in the sea. On the physical side, Matthew Fontaine Maury, developing and extending Franklin's earlier work, made comprehensive computations of wind and current data and set up the machinery for international cooperation. His book, *Physical Geography of the Sea*, has been regarded as the first text in oceanography.

Forbes and Maury were followed by a distinguished group of men whose interest in oceanography led them to make the first truly oceanographic expeditions. Most famous of these was the three-year around-the-world voyage of HMS *Challenger*, which followed earlier explorations of the *Lightning* and *Porcupine*. Instrumental in organizing these was Wyville Thompson, later joined by John Murray. Later in succession were the Norwegian Johan Hjort and the *Michael Sars* North Atlantic exploration; Louis Agassiz; Albert Honoré Charles, Prince of Monaco, in a series of privately owned yachts, named *Hirondelle* I and II and *Princess Alice* I and II. Other important contributions were made by Michael and G. O. Sars, Helland-Hansen, Carl Chum, Victor Hansen, Otto Petterson, Gustav Ekman, and the vessels *Valdivia*, the Danish *Dana*, the British *Discovery*, the German *National* and *Meteor*, and the Dutch *Ingold*, *Snellius*, and *Siboga*, the French *Travailleur* and *Talisman*, the Austrian *Pola*, and the North American *Blake*, *Bache*, and *Albatross*. Among the North American pioneers were Alexander Agassiz, L. F. de Pourtales, and J. D. Dana. Pioneers in modern oceanographic work are M. Kunelsen, Sven Ekman, A. S. Sverdrup, A. Defant, Georg Wüst, Gerhard Schott, and Henry Bigelow.

Modern oceanography relies less upon single explorations than upon the continuous operation of single vessels belonging to permanent institutions, such as *Atlantis* of the Woods Hole Oceanographic Institution, *Horizon* of Scripps Institution of Oceanography, *Vema* of the Lamont Geological Observatory, the French *Calypso*, and the large Russian vessels *Vitiaz* and *Mikhail Lomonosov*. Single explorations continue to be made, as exemplified by the Swedish *Albatross* and Danish *Galatea*.

The reduction of data and study of collections from earlier expeditions were carried out generally in research institutions, museums, and universities not solely or primarily engaged in oceanography. The first marine laboratories were interested principally in fishery problems or were designed as biological stations to accommodate visiting investigators. Many of the former have extended their activities to cover chemical and physical oceanography during their growth and development. The [illegible] extensions of university bio- [illegible] the Naples [illegible] ogical Labo- [illegible] ach stations [illegible] contribute greatly to the development of biology, generally in such fields as embryology and physiology.

The number of institutions devoted to organized oceanographic investigations with permanent scientific staffs has gradually grown. At first the requirements of fishery research provided the stimulus in countries adjacent to the North Sea, but in later years laboratories in other countries wholly or mainly devoted to oceanography have grown considerably in number. A few may be mentioned here. In England, among other important institutions are the National Institute of Oceanography, the Marine Biological Laboratory at Plymouth, and the Fisheries Laboratory at Lowestoft. In the United States are the Woods Hole Oceanographic Institution in Massachusetts, the Scripps Institution of Oceanography in California, the Lamont Geological Observatory in New York, the University of Miami Marine Laboratory in Florida, the Texas A. & M. College Department of Oceanography, and the Oceanography Laboratories of the University of Washington at Seattle. In Germany oceanographic laboratories are located at Kiel and at Hamburg. In Denmark the Danish Biological Station is at Copenhagen. Other European laboratories include those at Bergen, Norway; Göteborg and Stockholm, Sweden; Helsinki, Finland; and at Trieste. France. Laboratories are located at Toyko, Japan; Nanaimo and Halifax, Canada; and in Hawaii. This list is not inclusive and necessarily leaves out a considerable number of important institutions. [F.G.W.S.]

Oceanographic surveys. Oceanographic surveys require careful planning because of high cost. Provision must be made for the proper type of vessel, equipment, and laboratory facilities, adapted to the nature and duration of the survey.

Research ships. Ships of all types and sizes have been gathering information about the oceans since earliest times. Vessels of less than 300 tons displacement seldom range farther than several hundred miles from land, whereas ships larger than 300 tons displacement may work in the open ocean for several months at a time. Research ships of all sizes must be seaworthy and must provide good platforms from which to work. More specifically, a ship must have comfortable quarters, adequate laboratory and deck space for preliminary analyses, plus storage space for equipment, explosives, samples and scientific data. Machinery, usually in the form of winches and booms, is necessary for handling the complex and often heavy scientific equipment needed to probe the ocean depths (Fig. 2).

Standard oceanographic equipment includes collecting bottles (Nansen bottles), for obtaining water samples, and thermometers (both reversing thermometers and bathythermographs), for measuring temperatures at all depths. In addition, there are various devices for obtaining samples of ocean bottom sediments and biological specimens. These include heavy coring tubes which punch cylindrical sediment sections out of the bottom, dredges which scrape rock samples from submerged mountains and platforms, plankton nets for collecting very small planktonic organisms, and trawls for collecting larger free-swimming orga[illegible]ms at all [illegible]

problem is the composition of sea water and the extent to which the ratio among the components is constant In connection with these problems, it has been urged that a library of water samples be established Improved techniques of measuring sound velocity, electrical conductivity, refractive index, and density must be developed to enable scientists to follow many processes in the ocean It is therefore necessary to study the small variations of these parameters in the sea See SEA WATER

The tides in the oceans are rather well known at the surface but are almost completely unknown in the deep sea, also, the influence of land boundaries on deep sea tides is not yet understood Further research also must be devoted to the interesting phenomenon of internal waves See TIDE; WAVE (INTERNAL)

The study of ocean waves is one of the most advanced topics in oceanography but the energy exchange between atmosphere and sea surface by friction must be studied further Another problem is that of the heat exchange between ocean and atmosphere, an important link in the heat mechanism which determines the weather and the oceanic circulation See OCEAN WAVES; OCEAN-METEOROLOGICAL RELATIONS.

The climate of the past, in particular that of the last 1,000,000 years, is best studied in the ocean New isotopic methods in paleoclimatologic research allow the determination of temperature variations in the ocean with a high degree of accuracy The rapid growth of geochemistry and the increased sampling of deep-sea sediments through improved techniques have solved some of the problems of deep sea sedimentation At the same time a number of new ones have been created such as: Why is the sediment carpet only about 300 *m* thick? What is the ma ... hat is the ... What is ... DIMENTS, ... MARINE TOPOGRAPHY, see also GEOLOGIC THERMOMETRY; PALEOECOLOGY (GEOCHEMICAL ASPECTS)

Many new instruments for the study of the ocean floor have been developed Many more are under construction Drilling in the ocean floor may result in many new questions A few cores only 500 m long will reveal the history of the earth over several million years, and a core to the Moho (about 3 miles deep) may help answer many of the questions about the structure of the earth's crust The problem of the mechanism of formation of the ridges and island chains may be near solution See MOHO (MOHOROVIČIĆ DISCONTINUITY); MOHOLE.

The age determination of sediments by radioactivity methods, which was thought impossible 30 years ago, is now used on all deep-sea sediments but covers a period of only 400 000 years. Improved chemical methods using beryllium 10 may push the age limit back to 5,000,000 years. Also ... little is known ... the sea ... ments in ... rates in ...

processes on the sea floor will be an important factor as they seem to produce chemical energy in sediments. See HYDROSPHERE, GEOCHEMISTRY OF.

In marine biology, the systematics and ecology remain the major aspects It is still the science of the "naturalist." The interest of marine biology is many-sided and not grouped around a few central problems. Ocean life offers to the general biologist the best opportunities to study such complex problems as the structure of communities and the flux of energy through these communities. The zonation of animals on the shore and in the open ocean is not yet fully understood, the cause of patchiness in the distribution must be found On the other hand, the distribution of species by currents and eddies must be studied and large-scale experiments on behavior must be carried out. Observation at sea has been neglected to a large extent, and therefore the equilibrium between sea observation and laboratory experiment ... the sea. See DEEP-SEA FAUNA; MARINE ECOSYSTEM; MARINE MICROBIOLOGY. [F.F.K.]

Applications of ocean research. Directly and indirectly the ocean is of great importance to man. It is valuable as a reservoir of natural resources, an outlet for waste disposal, and a means of transport and ... face and submarine phenomena. In all of these aspects oceanography provides basic information for engineers who seek to increase its benefits and to avoid its harmful effects. See MARINE RESOURCES.

Food resources. The food resources of the ocean ... the source of living energy. Nevertheless, this potential is only in part realized. Oceanographic studies provide information which can help to increase fishing yields through improved exploratory fishing, economical harvesting methods, fisheries forecasts, processing techniques at sea, and aquaculture.

Fishes are dependent in their distribution upon food organisms and plankton, vertical and horizontal currents which bring nutrients to the ... bottom ... a ... t ... a ... c ... a ... measurements of the basic organic productivity of the sea may become essential in the selection of regions for extended fishery exploration See SEA WATER FERTILITY.

The catching of fishes may be facilitated and newer and more efficient methods devised through a knowledge of the reaction of fishes to stimuli, and of their habits in general This knowledge may re-

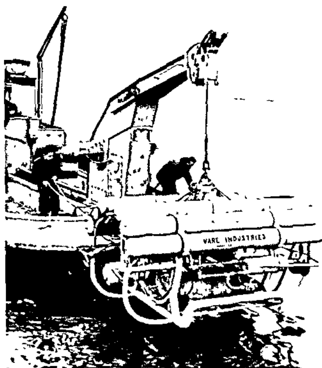


Fig. 4. This remote-controlled underwater television camera, mounted in self-propelled vehicle, can make visual surveys to depths of 1000 ft. The self-buoyant unit moves about or hovers at any desired depth in currents or tides of several knots. (*Vare Industries, Roselle, New Jersey*)

has now been developed for taking successive 2-m sediment cores to possible depths of 100 meters or more. See UNDERWATER PHOTOGRAPHY; UNDERWATER TELEVISION.

Oceanographic stations. Work on station consists of sampling and measuring as many marine properties as possible within the limitations of an expedition. Water, sea-bottom, and biological samples are successively collected on the long cables extended to the ocean floor. In some surveys, one cable lowering may include samplers for all these items, but this is not the usual procedure. Stations are systematically located at predetermined points along the ship's path. At hydrographic stations, observations of water temperature, salinity, oxygen, and phosphate content are determined upon sample recovery. Seismic stations generally are carried out by two ships, one ship running a fixed course and dropping explosives while the second ship remains stationary and records the returning subbottom reflected or refracted sound waves. Biological stations may consist of vertical net hauls or horizontal net tows depressed to sweep the ocean at a fixed depth. Geological stations are usually coring or bottom-dredging operations. Wherever possible, the recovered data are given a preliminary reduction aboard ship so that interesting discoveries are not [illegible] obtained. [illegible] possible [illegible] and laboratory equipment. Instead, the carefully processed, labeled, and stored material is preserved for intensive study ashore.

Home laboratory. This phase of the work may entail many months of careful examination and detailed analyses. Batteries of sophisticated scientific instruments are often necessary: data computers for reduction of physical oceanography information; spectrographic apparatus consisting of emission units; infrared, ultraviolet, x-ray, and mass spectrometers for chemistry; aquaria, pressure chambers, and chemostats for the biologist; electron microscopes and high-powered optical microscopes for examination of inorganic and organic constituents; radioisotope counters; and numerous standard physical, chemical, geological, and biological instruments. The great variety of measurable major and minor properties is reduced to statistical parameters which may then be integrated, correlated, and charted to increase the knowledge of sea properties and show their relationships. The essentially descriptive properties lead to an understanding of the principles which control the origin, form, and distribution of the observed phenomena. The present knowledge of the oceans is still fragmentary, but the increasing store of information already is being applied to a rapidly expanding number of man's everyday problems.

[G.A.R.]

Research problems in oceanography. Since the middle of the 19th century, man has learned more and more about that 71% of the earth which is covered with sea water. The rate of increase of knowledge is being accelerated by improved tools and methods, and increased interest of scientists and engineers. Some of the present and future research problems that are attracting scientists will be mentioned here.

One of the oldest and still unsolved problems i[s] the motion of ocean waters, involving surface cur[-]rents, deep-sea currents, vertical and horizont[al] turbulent motion, and general circulation. Ne[w] methods such as distribution of radioactive sub[-]stances, deep-sea current meters and neutrall[y] buoyant floats, high-precision determination of sal[t] and gas content, and hot-wire anemometers fo[r] turbulence studies and current measurements hav[e] increased present knowledge considerably. The sur[-]face movement of water is wind-produced, and [a] general theory of the motion has been worked ou[t]. The deep-sea currents, which are known to b[e] caused in part by variations in the thermohalin[e] circulation, are still an open problem. Superim[-]posed on these movements is the turbulent motio[n] which ranges over the whole spectrum from larg[e] ocean eddies transporting millions of cubic mete[rs] of water per second to the tiniest vibrations [of] water particles. Very little is known about turbu[-]lence. See OCEAN CURRENTS.

The mixing of water masses and the formation [of] new water masses cannot yet be completely an[d] adequately described, as many of the thermody[-]namic parameters are not precisely known. Labora[-]tory experiments and measurements of thermal ex[-]pansion, saline contraction, and specific heat [at] constant pressure must be carr[ied] out. A furth[er]

problem is the composition of sea water and the extent to which the ratio among the components is constant. In connection with these problems, it has been urged that a library of water samples be established Improved techniques of measuring sound velocity, electrical conductivity, refractive index, and density must be developed to enable scientists to follow many processes in the ocean. It is therefore necessary to study the small variations of these parameters in the sea *See* SEA WATER

The tides in the oceans are rather well known at the surface but are almost completely unknown in the deep sea; also, the influence of land boundaries on deep-sea tides is not yet understood. Further research also must be devoted to the interesting phenomenon of internal waves *See* TIDE; WAVE (INTERNAL)

The study of ocean waves is one of the most advanced topics in oceanography but the energy exchange between atmosphere and sea surface by friction must be studied further Another problem is that of the heat exchange between ocean and atmosphere ...

... LOGICAL RELATIONS

The climate of the past, in particular that of the last 1,000 000 years, is best studied in the ocean. New isotopic methods in paleoclimatologic research allow the determination of temperature variations in the ocean with a high degree of accuracy. The rapid growth of geochemistry and the increased sampling of deep-sea sediments through improved techniques have solved some of the problems of deep-sea sedimentation At the same time, a number of new ones have been created such as: Why is the sediment carpet only about 300 m thick? What is the mechanism of sediment transport? What is the history of sea water? Of the ocean basin? What is the cause of ice ages? *See* MARINE SEDIMENTS, SUBMARINE TOPOGRAPHY, *see also* GEOLOGIC THERMOMETRY, PALEOECOLOGY (GEOCHEMICAL ASPECTS)

Many new instruments for the study of the ocean floor have been developed Many more are under construction Drilling in the ocean floor may result in many new questions A few cores only 500 m long will reveal the history of the earth over several million years, and a core to the Moho (about 3 miles deep) may help answer many of the questions about the structure of the earth's crust The problem of the mechanism of formation of the ridges and island chains may be near solution *See* MOHO (MOHOROVICIC DISCONTINUITY); MOHOLE

The age determination of sediments by radioactivity methods, which was thought impossible 30 years ago, is now used on all deep-sea sediments but covers a period of only 400 000 years Improved chemical methods using beryllium-10 may push the age limit back to 5,000,000 years Also, very little is known about the formation of minerals on the sea floor the diffusion and adsorption of elements in and on sediments, and the reaction at slow rates in sediments. Certainly the microbiological processes on the sea floor will be an important factor as they seem to produce chemical energy in sediments *See* HYDROSPHERE, GEOCHEMISTRY OF.

In marine biology, the systematics and ecology remain the major aspects It is still the science of the "naturalist." The interest of marine biology is many-sided and not grouped around a few central problems. Ocean life offers to the general biologist the best opportunities to study such complex problems as the structure of communities and the flux of energy through these communities. The zonation of animals on the shore and in the open ocean is not yet fully understood; the cause of patchiness in the distribution must be found. On the other hand, the distribution of species by currents and eddies must be studied and large-scale experiments on behavior must be carried out. Observation at sea has been neglected to a large extent, and therefore the equilibrium between sea observation and laboratory experiment must be restored. The great advances in genetics, biochemistry, physiology, and microbiology also will advance the study of life in the sea *See* DEEP-SEA FAUNA; MARINE ECOSYSTEM; MARINE MICROBIOLOGY. [F.F.K.]

Applications of ocean research. Directly and indirectly the ocean is of great importance to man. It is valuable as a reservoir of natural resources, an outlet for waste disposal, and a means of transport and communication. The ocean is also important as a harmful agent causing biological, chemical, and mechanical destruction of life and property. In addition to the peaceful exploration of the oceans, there are many military applications both of surface and submarine phenomena. In all of these aspects oceanography provides basic information for engineers who seek to increase its benefits and to avoid its harmful effects. *See* MARINE RESOURCES.

Food resources. The food resources of the ocean are potentially greater than those of the land, since its larger area receives a proportionately larger amount of solar radiation, the source of living energy Nevertheless, this potential is only in part realized Oceanographic studies provide information which can help to increase fishing yields through improved exploratory fishing, economical harvesting methods, fisheries forecasts, processing techniques at sea, and aquaculture.

Fishes are dependent in their distribution upon food organisms ...

... to these environmental conditions and of the distribution of these conditions in the oceans is vital to successful extension of fishing areas Satisfactory measurements of the basic organic productivity of the sea may become essential in the selection of regions for extended fishery exploration. *See* SEA WATER FERTILITY.

The catching of fishes may be facilitated and newer and more efficient methods devised through a knowledge of the reaction of fishes to stimuli, and of their habits in general. This knowledge may re-

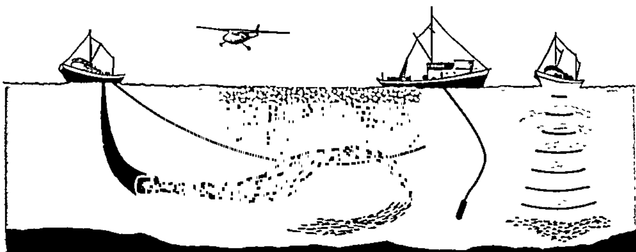


Fig. 5. Sketch showing new ways of fishing spotting by aircraft and sonar, electric and air bubble fence. (Adapted from A Spilhaus, Turn to the Sea, NAS-NRC, 1959)

sult in better design of nets and in the use of electrical, sonic, photic, and chemical traps or baits (Fig. 5). The harvest also may be increased by using improved methods of locating schools of fish by sonic or other means.

The biology of fishes, their food preferences, their predators, their relation to oceanographic conditions, and the fluctuations in these conditions seasonally and from year to year are important factors in forecasting fluctuations in the fisheries This information will aid in preventing the economic waste of alternating glut and scarcity Similar information is essential to good management of the fisheries and to sound regulation by conservation agencies

Other anticipated advances which require further oceanographic study include (1) the improvement of fishing by transplanting the young of existing stocks or by introducing new stocks; (2) farming or cultivation of sea fishes (although this does not seem feasible at present, scientific research has improved the cultivation of oysters and mussels in France and Japan); and (3) the use of planktonic vegetation as a source of food or animal nutrition *See* FISHERIES CONSERVATION; FOOD MANUFACTURING; MARINE FISHERIES.

Mineral resources. Although most of the valuable chemical elements in sea water are in very great dilution, the great volume of water of the oceans (about 300,000,000 mi^3) provides a limitless and readily accessible reservoir, if such dilute concentrations can be economically extracted. Magnesium is produced largely from sea water, and bromide also has been extracted commercially. High concentrations of manganese are found in manganese nodules which are very common on certain areas of the sea floor.

Other elements occur in too great dilution to be extracted by present methods. Possibly a better understanding of the ability of certain marine plants and animals to accumulate and concentrate elements from sea water in their tissues may lead the way to new methods of recovering these elements.

Energy and water source. Sea water contains deuterium and would be a limitless source for this element in the event of successful nuclear fusion developments. Further oceanographic knowledge and advances in engineering may lead to the increased utilization of tidal energy sources, or of the heat energy available from temperature differences in the ocean. The development of new methods for removing salt from sea water offers promise that the sea may become a practical source for potable water.

Disposal outlet. Because of the large volume of the sea, it is frequently used for the disposal of chemical wastes, sewage, and garbage. A knowledge of local currents and tides, as well as of the bottom fauna, is essential in order to avoid pollution of beaches or commercial fishing grounds. Radioactive waste disposal in offshore deeps poses problems of the rate of movement of deep waters and the transfer of radioactive materials through migration and food chains of marine organisms

Traffic and communication. The sea still remains an important highway; thus the knowledge and forecasting of waves, currents, tides, and weather in relation to navigation are of great practical importance New developments include the continuous rerouting of ships at sea in order that they may follow the most economic paths in the face of changing weather conditions

A knowledge of submarine topography, geologic processes, and temperature conditions is important for the satisfactory location, operation, and repair of submarine cables The use of Sofar in air-sea rescue operations is based upon submarine acoustics.

For additional information pertaining to overseas communication *see* TELEGRAPHY; TELEPHONE SERVICE

Defense requirements Defense aspects of marine research involve not only the navigation of surface vessels but also undersea craft with special navigational problems related to submarine topography, echo sounding, and the [illegible]tribution of t[illegible]perature, density, and oth[illegible]

in submarine acoustics has improved communication between, and detection of, undersea craft. In spite of these advances natural conditions, such as warm water pockets and subsurface magnetic irregularities, can conceal submarines from conventional means of detection. Investigation of these conditions is essential to any defense against missile carrying submarines See ANTISUBMARINE WARFARE.

Property and life. Damage to docks and ships by marine borers and fouling organisms is controlled by methods that utilize a knowledge of the biology, behavior and physiology of the destructive organisms, and of the oceanographic conditions which control their distribution (see BORING SPONGES; SHIPWORM; THORACICA). Loss of life caused by the attacks of sharks and other fishes may be reduced through an understanding of their behavior and the development of repellants and other protective devices. The chemical characteristics of sea water pose special problems of corrosion of metals. Beach erosion, wave damage to harbor and offshore structures, the effects of tsunamis and internal waves, and storms also cause loss of property and life. Much of this damage may be minimized by the application of oceanographic knowledge to forecasting methods and warning systems. See SHORE PROCESSES, STORM SURGE; TSUNAMI

Indirect benefits of oceanography arise from the application of marine meteorology to weather prediction, not only over the sea areas but also over the land (see MARINE INFLUENCE ON WEATHER AND CLIMATE) The study of marine geology and marine ecology aid in the understanding of the character of oil-bearing sedimentary rocks found on land

[F G.W S]

Bibliography H Barnes, *Oceanography and Marine Biology*, 1959. H B Bigelow, *Oceanography Its Scope, Problems, and Economic Importance*, 1931; G Dietrich and K Kalle, *Allgemeine Meereskunde*, 1959, W A Herdman, *Founders of Oceanography and Their Work, an Introduction to the Science of the Sea*, 1923, F G. W Smith and H Chapin, *The Sun, the Sea, and Tomorrow*, 1954, A F Spilhaus *Turn to the Sea*, 1959; H U Sverdrup, M W Johnson, and R H Fleming, *The Oceans*, 1942

Oceans and seas

The interconnecting body of salt water that covers 70.8% of the surface of the earth is called the world ocean, or simply the ocean Its major subdivisions, corresponding to the continents, are oceans

Subdivisions of oceans in turn are called seas, these range all the way from vague regions with no fixed limits (such as Sargasso Sea) to almost completely landlocked bodies (Black Sea). The terms bight strait, gulf, and bay are often used interchangeably with sea (Great Australian Bight, Denmark Strait, Gulf of Mexico, Bay of Bengal). Salt lakes lacking outlets to the ocean are also usually called seas (Salton Sea, Dead Sea, Caspian Sea).

Table 1. Characteristics of the oceans

Ocean	Area 10^3 m^3	Mean depth m	Volume 10^{15} m^3
Arctic	14,090	1,205	17 0
North Atlantic	46,772	3,285	153 6
South Atlantic	37,364	4,091	152 8
Indian	81,602	4,284	349 6
North Pacific	83,462	3,858	322 0
South Pacific	65,521	3,891	254 9
Antarctic	32,249	3,730	120 3

Table 2 Characteristics of individual seas

Sea	Area 10^3 m^3	Mean depth m	Volume 10^{12} m^3
Tributary to Arctic Ocean			
Norwegian Sea	1,383	1,742	2,408
Greenland Sea	1,205	1,444	1,740
Barents Sea	1,405	229	322
White Sea	90	89	8
Kara Sea	883	118	104
Laptev Sea	650	519	338
East Siberian Sea	901	58	53
Chukchi Sea	582	88	51
Beaufort Sea	476	1,004	478
Baffin Bay	689	861	593
Tributary to North Atlantic			
North Sea*	600	94	55
Baltic Sea*	386	86	33
Mediterranean Sea*	2,516	1,494	3,758
Black Sea*	461	1,166	537
Caribbean Sea*	2,754	2,491	6,860
Gulf of Mexico*	1,543	1,512	2,332
Gulf of St Lawrence	238	127	30
Hudson Bay	1,232	128	158
Tributary to South Atlantic			
Gulf of Guinea	1,533	2,996	4,592
Tributary to Indian Ocean			
Red Sea	450	558	251
Persian Gulf	241	40	10
Arabian Sea	3,863	2,734	10,561
Bay of Bengal	2,172	2,586	5,616
Andaman Sea	602	1,096	660
Great Australian Bight	484	950	459
Tributary to North Pacific			
Gulf of California	177	818	145
Gulf of Alaska	1,327	2,431	3,226
Bering Sea*	2,304	1,598	3,683
Okhotsk Sea	1,590	859	1,365
Japan Sea	978	1,752	1,713
Yellow Sea	417	40	17
East China Sea	752	349	263
Sulu Sea	420	1,139	478
Celebes Sea	472	3,291	1,553
In both North and South Pacific			
South China Sea	3,685	1,060	3,907
Makassar Strait	194	967	188
Molukka Sea	307	1,880	578
Ceram Sea	187	1,209	227
Tributary to South Pacific			
Java Sea	433	46	20
Bali Sea	119	411	49
Flores Sea	121	1,829	222
Savu Sea	105	1,701	178
Banda Sea	695	3,064	2,129
Ceram Sea	187	1,209	227
Timor Sea	615	406	250
Arafura Sea	1,037	197	204
Coral Sea	4,791	2,394	[illegible]

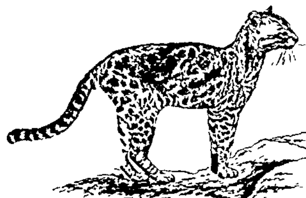
The ocean has a total area of 361×10^6 km^2 and an average depth of 3795 m, with a total volume of 1.37×10^{18} m^3. It has a mean temperature of 3.90°C and a mean specific gravity of 1.045, giving a total mass of 143×10^{16} metric tons. Since the mean salinity is 34.75‰ (3.475% by weight), there are 138×10^{16} tons of water and 4.87×10^{16} tons of salt in the ocean.

The world ocean, on the basis of its surface circulation, is conveniently divided into seven oceans (see Table 1). The Arctic Ocean is separated from the North Pacific at Bering Strait and from the North Atlantic by Fury and Hecla Strait, Davis Strait, and lines from Greenland to Iceland, Iceland to Scotland, and Scotland to Norway. The Equator divides the North Pacific from the South Pacific and the North Atlantic from the South Atlantic. The meridian of Cape Agulhas (20°E) divides the South Atlantic and Indian Ocean. The Indian and Pacific Oceans are separated by a line from Singapore to Sumatra; Indonesia; a line from Timor to Cape Talbot, Australia; the western end of Bass Strait; and the meridian of Southeast Cape, Tasmania (146°52'E). All of Magellan Strait is part of the Pacific. The Antarctic Ocean is all the water south of 55°S. Using these limits, the oceans and their adjacent seas have the characteristics listed in Tables 1 and 2.

See ANTARCTIC OCEAN; ARCTIC OCEAN, ATLANTIC OCEAN; INDIAN OCEAN; PACIFIC OCEAN; SOUTHEAST ASIAN WATERS. [J.LY]

Ocelot

A medium-sized American cat, *Felis pardalis*, found from Paraguay northward through Mexico and, rarely, in Texas and Arizona The ocelot may grow to 50 in. in length but is usually somewhat smaller, the tail comprises a third or more of its length. It is tawny gray above and marked with black or black-bordered spots, some of which are arranged in streaks.

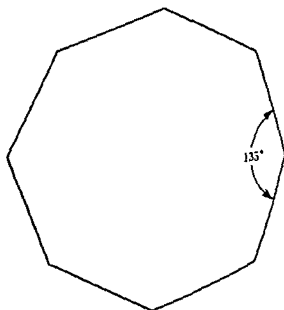


The ocelot, *Felis pardalis*; length 3 ft (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Ocelots are sometimes kept as pets, and the fur is of some value, especially as trimming for clothes. The animal is primarily nocturnal and prefers broken, wooded country. Although they usually catch their prey on the ground, ocelots are also capable of climbing trees. Almost any vertebrate of moderate size may be utilized as food by the ocelot. *See* CARNIVORA. [J.D.R.]

Octagon

A figure formed by the eight line segments (sides) that join in order eight ordered points (vertices) of a plane. In elementary geometry it is assumed that the sides do not cross and, usually, even that the finite region of the plane bounded by the sides is convex (convex octagon). A regular octagon has each two of its side congruent, and each angle made by adjacent sides equal to 135°. The area of a regular octagon of side a is $2(\sqrt{2}+1)a^2$. If a



Regular octagon

regular octagon be inscribed in a circle of un[it] radius, a rotation about the center of the circle wi[ll] make its vertices coincide with the eight (comple[x]) roots of the equation $z^8 - 1 = 0$, namely

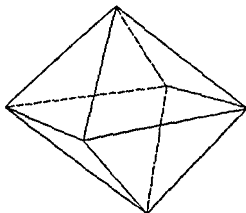
$$z_n = \cos(2\pi n/8) + i\sin(2\pi n/8)$$
$$n = 0, 1, 2, \ldots, 7$$

where $i = \sqrt{-1}$. *See* POLYGON; POLYTOPES, REGULAR [L.M.BL.]

Octahedron

In geometry, a solid with 8 faces. The faces of [a] regular octahedron (one of the 5 regular solids a[ll] of which were known to the ancient Greeks) a[re] all congruent equilateral triangles. These 8 tri[angles] have a total of 6 vertices and 12 edges—th[e] vertices and edges of the octahedron. The solid [is] dual to the cube, which has 8 vertices, 6 faces, a[nd] 12 edges.

The centers of the faces of a regular octahedro[n] are the vertices of a cube, and, conversely, th[e] centers of the faces of a cube are the vertices o[f] a regular octahedron. The vo[lume] of an octahedro[n]



Regular octahedron.

with edge a is $a^3\sqrt{2}/3$. *See* POLYHEDRON; POLYTOPES, REGULAR [L.M.BL.]

Octal number system

A system in which numbers are represented as linear combinations of powers of 8. Below is shown the relationship between the octal system and the decimal system.

Decimal notation	Octal notation
$12 = 8^1 + 4$	$= 14$
$31 = 3 \times 8^1 + 7$	$= 37$
$123 = 1 \times 8^2 + 7 \times 8^1 + 3$	$= 173$

The octal system has come into importance because of its value in computer applications *See* NUMBER SYSTEMS

Octane

An alkane having the *formula* C_8H_{18} The 18 possible *isomers, all* of which have been prepared, range in boiling point from 99.3°C (2,2,4-trimethylpentane) to 125 6°C (*n*-octane). One isomer, 2,2 3,3-tetramethylbutane, is a crystalline solid having a high melting point, 101°C, only 5 5°C lower than its boiling point

The *branched-chain* octanes are obtained commercially by the catalytic alkylation of isobutane with the *n*-butylenes or isobutylene They have high antiknock rating and are therefore valuable constituents of gasoline

The standard reference fuel (100 octane number) in determining the antiknock rating of gasolines is 2,2,4-trimethylpentane (often incorrectly termed *isooctane*) *It is* produced by the hydrogenation of the mixture of 2,4,4 trimethyl-1 pentene, and 2,4,4-*trimethyl* 2 pentene obtained by the polymerization of isobutylene. *See* ALKANE, GASOLINE; OCTANE NUMBER [L.S.]

Octane number

A number applied to fuel for spark-ignition engines to indicate the fuel's ability to resist combustion knock *To determine* octane number of a fuel sample, a specially designed engine is operated under specified conditions with the *given fuel* The compression ratio of the engine is adjusted to give a standard knock intensity as measured by an approved knockmeter.

Without changing the compression ratio, the engine is then operated on blends of isooctane (2,2,4-trimethylpentane, C_8H_{18}) which is a fuel highly resistant to knock, and normal heptane (C_7H_{16}), a fuel which knocks readily. When a blend is found that knocks under the conditions to which the engine was adjusted with the same intensity as did the fuel sample, the percentage by volume of isooctane in the mixture is the octane number of the fuel sample (*see* COMBUSTION KNOCK).

Fuels with octane numbers above 100 are usually rated by finding the number of milliliters of tetraethyllead required per gallon of isooctane to give the same resistance to detonation as the fuel sample. A method used with aviation fuel uses the ratio of engine indicated mean effective pressure (mep) at incipient knock for the fuel being rated to the mep at incipient knock for isooctane. *See* ANTIKNOCK AGENTS. [A.R.B.]

Octave

The interval between two sounds having a basic frequency ratio of 2:1; also, the interval in pitch *between two tones such* that one tone may be regarded as duplicating at the next higher pitch the basic musical import of the other tone Both these ideas stem from the musical interval of the eighth; in general, an octave is simply a *group of eight.* The first definition given is concerned with physical oscillations, and *in this sense,* the interval in octaves between any 2 frequencies is the logarithm to the base 2 of the frequency ratio The second definition deals with sensations Histori-

... is not necessarily the case (*see* PITCH) The upper tone of a pair delimiting the interval of an octave is itself *called* the octave; the organ pipe producing the upper tone is also called the octave. *See* MUSICAL ACOUSTICS; SCALE (MUSIC) [R.W.Y.]

Octode

An 8-electrode vacuum tube used as a *mixer tube in* radio receivers. Six of the electrodes ...

... achieves the desired effects of low electrostatic *coupling* between the signal and oscillator circuits, while at the same time, yielding good *plate-circuit* characteristics. *See* VACUUM TUBE. [K.R.S.]

Octopoda

An order of the dibranchiate cephalopods containing the octopus (see *illustration*), argonaut, blanket octopus, and others. They possess 8 arms equipped

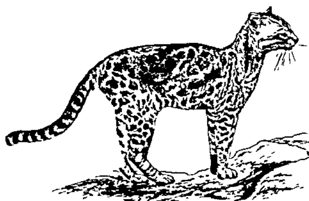
The ocean has a total area of 361×10^6 km² and an average depth of 3795 m, with a total volume of 1.37×10^{19} m³. It has a mean temperature of 3.90°C and a mean specific gravity of 1 045, giving a total mass of 143×10^{16} metric tons. Since the mean salinity is 34.75‰ (3 475% by weight), there are 138×10^{16} tons of water and $4 87 \times 10^{16}$ tons of salt in the ocean.

The world ocean, on the basis of its surface circulation, is conveniently divided into seven oceans (see Table 1). The Arctic Ocean is separated from the North Pacific at Bering Strait and from the North Atlantic by Fury and Hecla Strait, Davis Strait, and lines from Greenland to Iceland, Iceland to Scotland, and Scotland to Norway. The Equator divides the North Pacific from the South Pacific and the North Atlantic from the South Atlantic. The meridian of Cape Agulhas (20°E) divides the South Atlantic and Indian Ocean. The Indian and Pacific Oceans are separated by a line from Singapore to Sumatra; Indonesia; a line from Timor to Cape Talbot, Australia; the western end of Bass Strait; and the meridian of Southeast Cape, Tasmania (146°52′E). All of Magellan Strait is part of the Pacific. The Antarctic Ocean is all the water south of 55°S. Using these limits, the oceans and their adjacent seas have the characteristics listed in Tables 1 and 2

See ANTARCTIC OCEAN; ARCTIC OCEAN; ATLANTIC OCEAN; INDIAN OCEAN; PACIFIC OCEAN, SOUTHEAST ASIAN WATERS. [J LY]

Ocelot

A medium-sized American cat, *Felis pardalis*, found from Paraguay northward through Mexico and, rarely, in Texas and Arizona. The ocelot may grow to 50 in in length but is usually somewhat smaller; the tail comprises a third or more of its length. It is tawny gray above and marked with black or black-bordered spots, some of which are arranged in streaks.

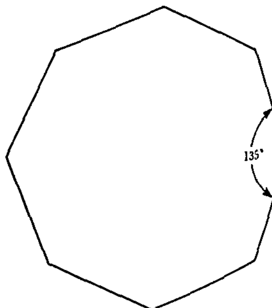


The ocelot, *Felis pardalis*, length 3 ft. (From E. L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Ocelots are sometimes kept as pets, and the fur is of some value, especially as trimming for clothes. The animal is primarily nocturnal and prefers broken, wooded country. Although they usually catch their prey on the ground, ocelots are al[so] capable of climbing trees. Almost any vertebrate [of] moderate size may be utilized as food by the ocel[ot]. *See* CARNIVORA. [J.D.]

Octagon

A figure formed by the eight line segments (side[s]) that join in order eight ordered points (vertice[s]) of a plane. In elementary geometry it is assum[ed] that the sides do not cross and, usually, even th[at] the finite region of the plane bounded by the si[des] is convex (convex octagon). A regular octagon [has] each two of its side congruent, and each an[gle] made by adjacent sides equal to 135°. The area o[f a] regular octagon of side a is $2(\sqrt{2}+1)a^2$ I[f a]



Regular octagon

regular octagon be inscribed in a circle of radius, a rotation about the center of the circle make its vertices coincide with the eight (compl[ex]) roots of the equation $z^8 - 1 = 0$; namely

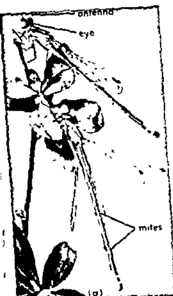
$$z_n = \cos(2\pi n/8) + i\sin(2\pi n/8)$$
$$n = 0, 1, 2, \ldots, 7$$

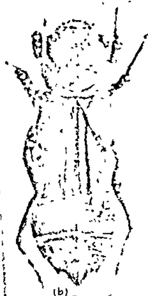
where $i = \sqrt{-1}$. *See* POLYGON, POLYTOPES, R[EGU]LAR. [L.M]

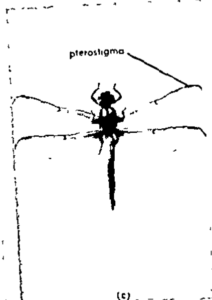
Octahedron

In geometry, a solid with 8 faces The faces [of a] regular octahedron (one of the 5 regular solid[s] of which were known to the ancient Greeks) [are] all congruent equilateral triangles. These 8 [tri]angles have a total of 6 vertices and 12 edges— vertices and edges of the octahedron. The sol[id is] dual to the cube, which has 8 vertices, 6 faces, 12 edges.

The centers of the faces of a regular octahed[ron] are the vertices of a cube, and, conversely, [the] centers of the faces of a cube are the vertice[s of] a regular octahedron. The vo[lume] of an octah[edron]







Typical Odonata, (a) Adult damselflies, note mites attached to these (b) Dragonfly nymph (c) Adult dragonfly

Dragonflies constitute one of the oldest insect orders; they can be traced back, through fossil records, to the Carboniferous and Permian. Surprisingly, few changes have occurred since then, but the order has become diversified in late years and has specialized to counter competition and to avoid enemies Wing venation is more complex and body structure more varied

Life history, habits, and enemies. There are three general stages in the life history, the egg, the nymph, sometimes called naiad, and the adult Development is usually slow, often requiring three to five years. Rarely is there more than one generation a year in the northern range. Adults may live for an extended period in summer Eggs are laid by insertion into plant stems, either beneath the water or just above the surface Others are dropped directly into the water and sink to the bottom where they hatch and the nymphs develop Flight periods may extend over the entire summer, or be limited to spring and early summer Mass migration flights have been observed, similar to those of birds and certain butterflies Individual species in North America may range from coast to coast or from the northern to southern boundary of the United States

Enemies include birds and fishes, with frogs and insects being of lesser importance Mites sometimes attach themselves (as may be seen in the illustration of the damsel fly) but actually do little harm

Classification. The Odonata comprise a relatively small order of insects with 112 species listed from Connecticut, 164 from northeast United States, and about 360 from the United States as a whole There are at least 500 from North America and probably less than 3000 species known throughout the world. The order is divided into the Anisoptera, or true dragonflies, and the Zygoptera, or damsel-

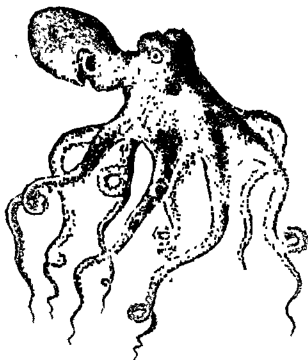
Anisoptera, or true dragonflies, are the more numerous of the two suborders. Adults have large wings and a thickset thorax Body colors are mostly browns, blacks, and blues, and a few have metallic lusters. Wings may be either clear or spotted Nymphs possess internal rectal gills for breathing. They inhabit the bottoms of lakes, streams, or ponds. They are squat and usually

the Gomphidae, Petaluridae, Aeshnidae, Cordulegasteridae, and Libellulidae, the largest.

Zygoptera, or damselflies, are slender, dainty creatures, often with bright blue or orange coloring and usually with clear or transparent wings. The nymphs are provided with long, thin, often transparent, gills attached to the tip of the abdomen and provided with tracheae, which enable the insect to breathe under water Nymphs are found among floating vegetation, or clinging to water plants where, as in the Anisoptera, their color and form protect them. Locomotion is by sculling with their tracheal gills or by crawling, as in the Anisoptera Adults often congregate in sunny spots along the shores or streams where they feed on small flies The principal families are Agrionidae, Lestidae, and Coenagrionidae; the last contains the greatest number of species.

[P C.]

Bibliography: J. G Needham and H. B Heywood, *A Handbook of the Dragonflies of North America*, 1929. J. G. Needh...



The octopus, *Octopus vulgaris*. (From P. Martin Duncan, ed., *Cassell's Natural History*, Cassell)

with 1–3 rows of suckers. The common *Octopus vulgaris* and its allies are secretive bottom dwellers in shallow to moderate depths. They feed upon bivalve mollusks and crustaceans and are often serious predators on lobsters and crabs Others, such as *Argonauta argo* with its fragile egg case, live in the open ocean. Many are small gelatinous forms occurring down to 3000 meters while larger deep sea forms may have feeble paddle-shaped fins and one, *Cirrothauma*, is blind *Paleoctopus newboldi* is from the Upper Cretaceous. The largest octopus recorded had an arm spread of 32 ft but most are much smaller. There are about 150 species *See* CEPHALOPODA; DIBRANCHIA. [G L.V]

Bibliography · G. C. Robson, *A Monograph of the Recent Cephalopoda*, vols. 1 and 2, 1929–1932

Octopus

Any of about 50 living species of the family Octopodidae, class Cephalopoda, phylum Mollusca. The octopuses, also known as devil fishes, are worldwide in distribution, are all marine, and are found in both deep and shallow water. There are three species on the Pacific Coast of the United States and four on the Atlantic Coast.

Uses. The octopuses are used for food in many parts of the world They are especially favored in Italy and Greece, China and other Oriental countries, and by the natives of the South Pacific islands. They are also of some importance as sport animals in many parts of the world, for example, in certain Pacific islands where the principal sport of teenage boys is diving for and capturing octopuses without the use of weapons of any kind In some areas they are highly destructive to crab and lobster populations. Although they are not nearly as dangerous as popular accounts indicate, they will attack man and the larger species are capable of killing humans.

Characteristics. Members of the family vary in size from those with an arm spread of 2 in to monsters 28 ft from arm tip to arm tip Many reports of gigantic specimens undoubtedly are misidentified observations of the giant squid

The typical octopus has a round body with a very large head and large eyes There are eight arms which are more or less webbed and alike except the third right arm of males, which is enlarged and modified as a copulatory organ Each arm bears two rows of sessile suckers The mantle is attached to the head by a broad, dorsal commissure. There are no fins. Octopuses are all adept at changing color and are consequently easily overlooked when not moving about

Octopuses are all nocturnal, carnivorous animals Crabs appear to be their favorite food, but other crustaceans, mollusks, and fishes are readily consumed. They have strong, beaked jaws with which they kill their prey

Reproduction. Sperm is transferred in spermatophores, or bundles, to the mantle skirt of the female by the modified copulatory arm of the male Eggs are deposited on the roof of the female's hiding place in ropelike strings or grapelike clusters and are guarded by the female until they hatch In some species hatching requires 6–8 weeks

Octopuses can walk skillfully and rapidly along the bottom by the use of their arms. They also swim backwards with their arms trailing by ejecting water through the siphon, in much the same manner as squids. Also, like the squids, they have an ink sac which opens into the anus and by means of which they can cloud the water to escape their enemies. In internal anatomy they are similar to squids.

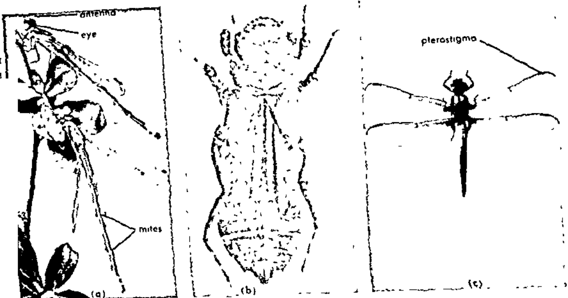
The most common Pacific Coast species is *Octopus bimaculatus* which ranges from Alaska to Lower California Although common in some localities, it has been so severely hunted in recent years for food that its numbers are somewhat reduced Its body is about 4 in long and is usually gray, though its color is highly variable. It is marked with two large round spots on the back.

Octopus bairdi, the most common Atlantic Coast species, has a 3-in. body, but its arms may attain a spread of 40 in It is normally bluish white speckled with brown. It is found in deep water from Cape Cod northward in North America, and Europe. *See* CEPHALOPODA, SQUID [J D

Odonata

An order of the class Insecta known as dragonflies The young inhabit ponds, streams, and marshes; the adults fly over these localities or adjacent la The adult structure is unique, characterized by head with large compound eyes and wings w clear or transparent membranes traversed by r works of veins, the male has accessory male ge tal organs possessed by no other insects.

Superstitions regarding dragonflies have p sisted for a long time Because of their bizarre pearance they have l ' devil's needles, snake doctors



Typical Odonata: (a) Adult damselflies, note mites attached to these (b) Dragonfly nymph (c) Adult dragonfly.

Dragonflies constitute one of the oldest insect orders; they can be traced back, through fossil records, to the Carboniferous and Permian Surprisingly, few changes have occurred since then, but the order has become diversified in late years and has specialized to counter competition and to avoid enemies Wing venation is more complex and body structure more varied

Life history, habits, and enemies. There are three general stages in the life history, the egg, the nymph, sometimes called naiad, and the adult Development is usually slow, often requiring three to five years Rarely is there more than one generation a year in the northern range. Adults may live for an extended period in summer Eggs are laid by insertion into plant stems, either beneath the water or just above the surface Others are dropped directly into the water and sink to the bottom where they hatch and the nymphs develop Flight periods may extend over the entire summer, or be limited to spring and early summer. Mass migration flights have been observed, similar to those of birds and certain butterflies Individual species in North America may range from coast to coast or from the northern to southern boundary of the United States

Enemies include birds and fishes, with frogs and insects being of lesser importance Mites sometimes attach themselves (as may be seen in the illustration of the damsel fly) but actually do little harm

Classification. The Odonata comprise a relatively small ...

There are at least 500 from North America and probably less than 3000 species known throughout the world The order is divided into the Anisoptera, or true dragonflies, and the Zygoptera, or damselflies.

Anisoptera, or true dragonflies, are the more numerous of the two suborders. Adults have large wings and a thickset thorax Body colors are mostly browns, blacks, and blues, and a few have metallic lusters Wings may be either clear or spotted Nymphs possess internal rectal gills for breathing. They inhabit the bottoms of lakes, streams, or ponds. They are squat and usually sluggish insects, protected by their form and color, which resemble objects on the bottom Locomotion is by a sort of jet propulsion; but they crawl slowly in stalking prey The principal families comprise the Gomphidae, Petaluridae, Aeshnidae, Cordulegasteridae, and Libellulidae, the largest.

Zygoptera, or damselflies, are slender, dainty creatures, often with bright blue or orange coloring and usually with clear or transparent wings. The nymphs are provided with long, thin, often transparent, gills attached to the tip of the abdomen and provided with tracheae, which enable the insect to breathe under water. Nymphs are found among *floating vegetation*, or clinging to water plants where, as in the Anisoptera, their color and form protect them. Locomotion is by sculling with their tracheal gills or by crawling, as in the Anisoptera Adults often congregate in sunny spots along the shores or streams where they feed on small flies. The principal families are Agrionidae, Lestidae, and Coenagrionidae, the last contains the greatest number of species

[P.C.]

Bibliography: J. G. Needham and H B. Heywood, *A Handbook of the Dragonflies of North America* 1929; ...
... *Canada and Alaska*, vol 1, 1953

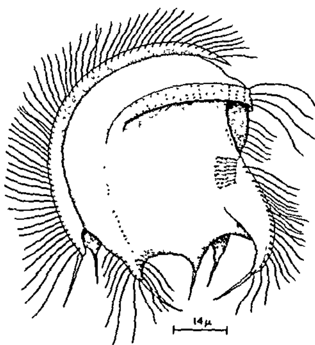
Odontognathae

One of the three superorders comprising the subclass Neornithes, or true birds. It is now considered to contain *only a single* order, Hesperornithiformes, of which the best-known family is the Hesperornithidae. The five species of this family are known only from Upper Cretaceous fossils from Kansas and Montana.

They were large, flightless, aquatic birds, with the shoulder girdle much reduced and the legs powerfully developed for swimming. The dentary and maxillary bones of the jaws bore well-developed teeth in grooves The Odontognathae were probably an evolutionary offshoot, not ancestral to any living birds. The order Ichthyornithiformes, formerly included here, has recently been removed and made the type of a separate superorder, Ichthyornithes *See* AVES; ICHTHYORNITHES; NEORNITHES. [K.C.P.]

Odontostomatida

An order of the Spirotricha which represents a minor group of small, bizarre-looking species. The odontostomes are compressed laterally and possess very little ciliature. Even the adoral zone of membranelles is reduced in prominence. These ciliates are found in sewage disposal environs and other fresh or salt water habitats which have a very low oxygen content. *Epalxis* and *Saprodinium* (see illustration) are commonly encountered species. The ordinal name for this group commonly, but inadvisedly, used in the literature is Ctenostomatida, a name preoccupied by a bryozoan group of invertebrates. *See* SPIROTRICHA. [J.O.C.]



Saprodinium, an example of an odontostomatid.

Oegophiurida

An extinct order of Ophiuroidea, with seven Paleozoic families, characterized by ambulacral ossicles fused in pairs to form vertebrae as in extant ophiuroids, but with an open ambulacral groove as in asteroids and the stenurid ophiuroids. *See* OPHIUROIDEA; STENURIDA. [H.B.F.]

Oersted

A unit of magnetic field strength in the centimeter-gram-second (cgs) electromagnetic system of units. An oersted is the field strength at the center of a plane circular coil of one turn and 1 cm radius when there is a current of $1/2\pi$ abamp in the coil. The relation of the oersted to the mks unit of magnetic field strength is found from the equation for the magnetic field strength H at the center of a flat circular coil of N turns and radius r, carrying a current I.

$$H = \frac{NI}{2r}$$

Then

$$1 \text{ oersted} = \frac{(1/2\pi) \text{ abamp} \times 1 \text{ turn}}{2 \times 1 \text{ cm}}$$

$$= \frac{10 \text{ amp} \times 1 \text{ turn}}{4\pi \times 0.01 \text{ m}} = \frac{10^3}{4\pi} \frac{\text{amp-turn}}{\text{m}}$$

or

$$1 \text{ amp-turn/m} = 4\pi \times 10^{-3} \text{ oersted}$$

The oersted is also defined from the force on a cgs unit magnetic pole Since H = force/pole, an oersted is a dyne per unit pole. *See* ELECTRICAL UNITS; MAGNETIC FIELD [K.V.M.]

Ohm

The unit of electrical resistance in the rationalized meter-kilogram-second (mks) system of units. When the applied potential difference is 1 volt, the current in a 1-ohm resistor is 1 ampere. The voltage drop across a resistor is given by the product of its resistance in ohms and the current in amperes This is often called the IR drop

Formerly, 1 ohm (International) was defined as the resistance of a column of mercury of uniform cross section of length 106.300 cm and mass 14.4521 grams at a temperature of 0°C. Since 1948 the legal standard has been the absolute ohm, which is most conveniently defined in terms of the wave impedance of empty space, although in practice the absolute ohm is determined by the reactance offered to alternating current by an inductor. *See* ELECTRICAL STANDARDS

In empty space the wave impedance Z_0 for electromagnetic waves is defined as the ratio of the amplitude of the electric field vector E to the amplitude of the magnetic field vector H. In the rationalized mks system the units of E are volts per meter, and of H ampere turns per meter, so that Z_0 has units of volts per ampere, or ohms. Its numerical magnitude is 376 [illegible] lute ohms

From electromagnetic theory, one finds that $Z_0 = \sqrt{\mu_0/\epsilon_0}$, where μ_0 is the permeability of free space (12.57×10^{-7} mks units) and ϵ_0 is the permittivity of free space (8.84×10^{-12} mks units). Furthermore, in any system of units, $\mu_0\epsilon_0 = 1/c^2$ where c is the velocity of propagation of electromagnetic waves (approximately 3×10^8 m/sec). Thus $Z_0 = \mu_0 c$. Since the velocity of electromagnetic waves is well known, assignment of a value to μ_0 suffices to determine the absolute ohm. See ELECTRICAL UNITS; ELECTROMAGNETIC RADIATION; RESISTANCE, ELECTRICAL. [J W ST]

Ohm's law

A physical law which states that, for a given circuit element, the ratio of voltage V to current I is constant for a range of values of V and I. If Ohm's law is obeyed, the potential drop V across the circuit element is directly proportional to the current I which is flowing in the element. The resistance R is a constant. Ordinary metallic conductors obey Ohm's law

tic

...ves as the definition of resistance. See RESISTANCE, ELECTRICAL.

The power dissipation P in any circuit element is given by $P = VI$, where P is in watts, V in volts, and I in amperes. The power can often also be expressed as I^2R or V^2/R (see JOULE'S LAW). These last two relations are not particularly useful if R is a function of I, as is the case when Ohm's law does not apply.

The major instances of departures from Ohm's law are as follows:

1. When a material is heated sufficiently by a current, its resistance is changed, as in the case of the filament of an incandescent electric light bulb. The resistance of a tungsten filament increases with current, so that the current increases less rapidly than linearly with the applied voltage. For semiconducting materials heated to high temperature the resistance decreases, and the current increases more rapidly than linearly with the applied voltage.

2. The alternating-current resistance of many types of circuit elements is a function of both current and frequency.

3. In a substance such as an ionized gas, where the number of available charge carriers increases with the potential difference, the current rises faster than linearly with the applied voltage. The effective resistance thus decreases with increasing current.

4. In a triode vacuum tube the plate current is a function of the grid bias. Under certain conditions plate current can be made to increase without any appreciable change in the applied plate voltage. The effective resistance of the tube is then nearly zero.

5. In semiconductors the available charge carriers can be used up so that further increases in applied voltage do not produce corresponding changes in current.

Such non-ohmic elements as those described in 3, 4, and 5 have important applications in electronics. [J W.ST.]

Ohmmeter

A small, portable instrument using a microammeter and associated circuitry to measure resistance by the voltmeter-ammeter method (see RESISTANCE MEASUREMENT). Additional circuits are usually included to measure ac and dc volts and amperes, and the instrument is called a volt-ohm-milliammeter, or multimeter (see Fig. 1).

A typical resistance-measuring circuit is shown in Fig. 2a. Figure 2b shows a simplified schematic diagram of the ×100 range of the circuit. In operation, the instrument is first adjusted for full-scale deflection of the meter with the measuring leads shorted. When the unknown resistance is added to the circuit, the current through the meter decreases according to the relation

$$R_x = \left[\frac{I_{RO}}{I_{RX}} - 1\right]\left[\frac{R_M R_s}{R_M + R_s} + R_p\right] \qquad (1)$$

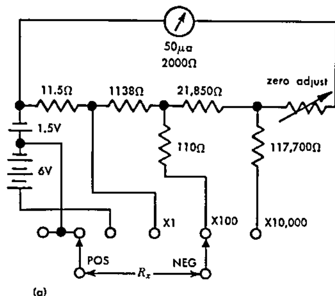
The various resistances are identified in Fig. 2b. I_{RO} is the current required for full-scale deflection of the meter when R_x is 0, and I_{RX} is the meter current with the unknown resistor R_x in the circuit. The meter deflects according to the value of I_{RX}.

Since R_x is a function of $1/I_{RX}$, the scale, calibrated in resistance units, is a reciprocal function. Full-scale deflection of the meter always corresponds to zero ohms, and no deflection indicates an open circuit.

The scale relation between these two points is arbitrary and is governed by the choice of one other scale point, such as the resistance for half-scale current. The scale of the instrument in Fig



Fig. 1 Volt-ohm-milliammeter. (Simpson Electric Co.)



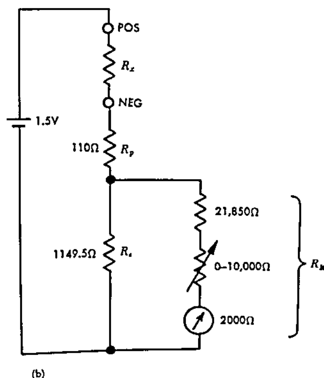


Fig. 2. Ohmmeter circuits. (a) Basic circuit (b) Simplified circuit of the ×100 range

1 is based on a half-scale resistance of about 12 ohms. In common with any measuring equipment having an error stated in terms of full-scale deflection, the absolute error (error of the reading) is large for small deflections.

Voltage and current ranges for this equipment are obtained in the usual fashion with resistance multipliers for extended voltage ranges, current shunts for multiple current ranges, and a copper-oxide rectifier for ac operation. A vacuum-tube voltmeter (VTVM) usually incorporates a resistance-measuring circuit for convenience. *See* AMMETER; VACUUM-TUBE VOLTMETER; VOLTMETER

[C.F.A.]

Oil analysis

The analysis of petroleum, and products derived from petroleum. Oil analysis falls into two categories, testing to classify the crude oil or product adequately for commercial uses, and analysis in terms of chemical composition.

Testing for the purpose of commercial transactions and specifications is based primarily on relatively simple tests, most of which have been cooperatively tested and issued as standard test methods by the American Society for Testing Materials. This organization publishes annually a manual, *ASTM Standards on Petroleum Products and Lubricants*, which contains detailed instructions for carrying out most tests used in the petroleum industry. A discussion of the significance of these tests is covered in a separate publication *Significance of ASTM Tests for Petroleum Products.* These two references should be consulted for the most up-to-date information on tests used in the petroleum industry.

The number and variety of commercial tests and specifications are so large that it is impractical even to list them here. The *ASTM Standards on Petroleum Products and Lubricants*, for example, runs to 1000 pages.

Petroleum consists primarily of compounds of carbon and hydrogen containing from 1 to about 60 carbon atoms. Roughly, the following division applies: C_1 and C_2 gas, C_3 and C_4 liquefied petroleum gas, C_4–C_{12} gasoline, C_{10}–C_{20} kerosine and light gas oil, and C_{20}–C_{40} lubricating oil and wax. Carbon atoms in natural petroleum occur in straight and branched chains (paraffins), single or multiple saturated rings (cycloparaffins or naphthenes), and in cyclic structures of the aromatic type such as benzene, naphthalene, and phenanthrene. Cyclic structures may have attached to them side chains of paraffinic carbons. In lubricating oil it is usual to have naphthene rings built onto the aromatic rings and side chains attached. In products which have been produced by cracking in the refinery olefins or compounds with carbon-carbon double bonds not in aromatic rings are also found.

The heavy fractions of petroleum contain increasing amounts of oxygen, nitrogen, and sulfur compounds as well as traces of organic compounds of metals such as vanadium. Asphalt contains a substantial proportion of oxygen. Crude oil usually contains some suspended or emulsified water and inorganic salts.

The analysis of crude petroleum or commercial products derived from petroleum in terms of chemical composition can be done with varying degrees of thoroughness. Anything approaching the complete analysis of a crude oil is so time-consuming and expensive that in the whole world only one sample of crude oil is being analyzed thoroughly. This project, known as Project No. 6, of the American Petroleum Institute, has been in progress for about 30 years, has cost over $1,500,000, and is far from

complete. There are, however, many methods of analysis which can be applied to crude petroleum and its products which give a reasonable amount of information about its chemical composition with relatively little effort

Crude-oil analysis. The primary steps in the analysis of crude oil are to determine its specific gravity (by hydrometer), the percentage of sulfur (by bomb), and the distribution of materials according to boiling range, as determined by distillation For example, material boiling between roughly 40 and 400°F is in the gasoline boiling range. Material from about 325 to 550°F is in the kerosine range Heavier fractions may be used for fuel oil, lubricating oil, wax, asphalt manufacture, or cracking to make more gasoline and fuel oil

The modern technical evaluation of crude oil by a purchaser depends upon many factors which are specific to each refinery. A complete evaluation requires much information on the composition and properties of the fractions of the crude obtained on distillation and a knowledge of the probable corrosive effects of the crude.

Whereas 20 years ago crude oils were bought and sold primarily on the basis of specific gravity, percentage of sulfur, and distillation data, the modern trend is to examine selected fractions by special techniques such as mass spectrometry, infrared spectroscopy, ultraviolet spectroscopy, gas chromatography, or physical property correlations to determine the suitability for reforming, cracking, and lubricating-oil production Each large purchaser of crude oil has developed a particular system. Although the methods used are far from uniform, the objective is in each case to obtain sufficient information about the composition of the crude oil to predict how the fractions will behave in modern refinery operations *See* PETROLEUM.

Analysis for composition. Hydrocarbon mixtures lighter than C_6 are analyzed for composition by low [illegible]

[illegible] from C_6 to C_8 are separated by gas-liquid chromatography which identifies most of the species Additional information is obtained by spectroscopic examination of any ambiguous cuts separated by chromatography

Heavier fractions require much more elaborate procedures although the distribution of hydrocarbon types within a narrow range of molecular weight can be approximated by fractional distillation or chromatography followed by spectroscopic analysis of the fractions

Olefins, aromatics, and saturates in gasoline and jet fuels can be separated by percolating through a tube of silica gel using fluorescent dyes to mark the boundaries This is the FIA analysis

The mass spectrometer is especially useful for the determination of naphthenes in saturated fractions because naphthenes with two rings can be distinguished from those with one ring

In very heavy oils, little can be done to identify specific hydrocarbons. Such information as the total saturates (separated by silica gel) and the concentration of various aromatic types (determined by spectroscopy) can be obtained

Carbon-type composition is so closely connected with the physical properties of oils that it can be derived from such properties by a variety of correlations which are accurate, over their range of usefulness, to about one-half of one carbon atom. The correlation based on density, refractivity intercept, and number of carbon atoms can be applied over the widest range of compositions.

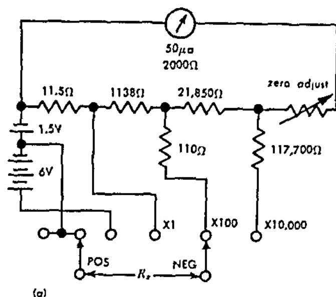
Analysis for sulfur, nitrogen, and some other inorganic elements can be made on all fractions by well known methods *See* PETROLEUM PROCESSING. [S.S K.]

Bibliography: American Society for Testing Materials (ASTM), *ASTM Standards on Petroleum Products and Lubricants*, 1959; ASTM, *Significance of ASTM Tests for Petroleum Products*, 3d ed., 1956; ASTM, *Composition of Petroleum Oils*, ASTM Spec. Tech. Pub. 224, 1958; F. D. Rossini, B. J Mair, and A J. Streif, *Hydrocarbons from Petroleum*, 1953; K. Van Nes and H. A. Van Westen, [illegible]
[illegible]
[illegible]
[illegible]

Oil and gas field development

The field development for petroleum. In the petroleum industry, a field means an area underlain without substantial interruption by one or more reservoirs of commercially valuable oil or gas, or both. A single reservoir (or group of reservoirs which cannot be separately produced) is a pool. Several pools separated from one another by barren, impermeable rock may be superimposed one above another within the same field. Pools have variable areal extent. Any sufficiently deep well located within the field should produce from one or more pools However, each well cannot produce from every pool, because different pools have different areal limits Development of a field includes the location, drilling, completion, and equipment of wells necessary to produce the commercially recoverable oil and gas in the field.

Related oil field conditions. Petroleum is a generic term which, in its broadest meaning, includes all naturally occurring hydrocarbons, whether gaseous, liquid, or solid By variation of the temperature, pressure, or both, of any hydrocarbon, it becomes gaseous, liquid, or solid Temperatures in producing horizons vary from 60° to more than 300°F, depending chiefly upon the depth of the horizon. A rough approximation is that temperature in the reservoir sands, or pay, equals 60°F plus 0 017°F/ft of depth below surface. Pressure on the hydrocarbons varies from atmospheric to more than 11,000 psi. Normal pressure is considered as 0 465 psi/ft of depth. Tem-



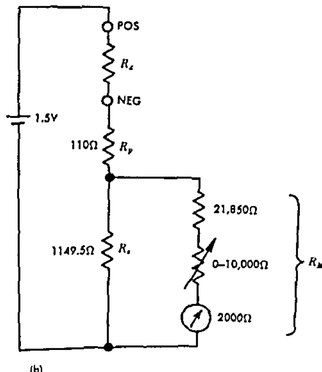


Fig. 2. Ohmmeter circuits. (a) Basic circuit. (b) Simplified circuit of the ×100 range.

1 is based on a half-scale resistance of about 12 ohms. In common with any measuring equipment having an error stated in terms of full-scale deflection, the absolute error (error of the reading) is large for small deflections.

Voltage and current ranges for this equipment are obtained in the usual fashion with resistance multipliers for extended voltage ranges, current shunts for multiple current ranges, and a copper-oxide rectifier for ac operation. A vacuum-tube voltmeter (VTVM) usually incorporates a resistance-measuring circuit for convenience. *See* AMMETER; VACUUM-TUBE VOLTMETER; VOLTMETER.
[C.E.A.]

Oil analysis

The analysis of petroleum, and products derived from petroleum. Oil analysis falls into two categories, testing to classify the crude oil or product adequately for commercial uses, and analysis in terms of chemical composition.

Testing for the purpose of commercial transactions and specifications is based primarily on relatively simple tests, most of which have been cooperatively tested and issued as standard test methods by the American Society for Testing Materials. This organization publishes annually a manual, *ASTM Standards on Petroleum Products and Lubricants,* which contains detailed instructions for carrying out most tests used in the petroleum industry. A discussion of the significance of these tests is covered in a separate publication *Significance of ASTM Tests for Petroleum Products.* These two references should be consulted for the most up-to-date information on tests used in the petroleum industry.

The number and variety of commercial tests and specifications are so large that it is impractical even to list them here. The *ASTM Standards on Petroleum Products and Lubricants,* for example, runs to 1000 pages.

Petroleum consists primarily of compounds of carbon and hydrogen containing from 1 to about 60 carbon atoms. Roughly, the following division applies: C_1 and C_2 gas, C_3 and C_4 liquefied petroleum gas, C_4–C_{12} gasoline, C_{10}–C_{20} kerosine and light gas oil, and C_{20}–C_{40} lubricating oil and wax. Carbon atoms in natural petroleum occur in straight and branched chains (paraffins), single or multiple saturated rings (cycloparaffins or naphthenes), and in cyclic structures of the aromatic type such as benzene, naphthalene, and phenanthrene. Cyclic structures may have attached to them side chains of paraffinic carbons. In lubricating oil, it is usual to have naphthene rings built onto the aromatic rings and side chains attached. In products which have been produced by cracking in the refinery, olefins or compounds with carbon-carbon double bonds not in aromatic rings are also found.

The heavy fractions of petroleum contain increasing amounts of oxygen, nitrogen, and sulfur compounds as well as traces of organic compounds of metals such as vanadium. Asphalt contains a substantial proportion of oxygen. Crude oil usually contains some suspended or emulsified water and inorganic salts.

The analysis of crude petroleum or commercial products derived from petroleum in terms of chemical composition can be done with varying degrees of thoroughness. Anything approaching the complete analysis of a crude oil is so time-consuming and expensive that in the whole world only one sample of crude oil is being analyzed thoroughly. This project, known as Project No. 6, of the American Petroleum Institute, has been in progress for about 30 years, has cost over $1,500,000, and is far from

complete. There are, however, many methods of analysis which can be applied to crude petroleum and its products which give a reasonable amount of information about its chemical composition with relatively little effort.

Crude-oil analysis. The primary steps in the analysis of crude oil are to determine its specific gravity (by hydrometer), the percentage of sulfur (by bomb), and the distribution of materials according to boiling range, as determined by distillation. For example, material boiling between roughly 40 and 400°F is in the gasoline boiling range Material from about 325 to 550°F is in the kerosine range. Heavier fractions may be used for fuel oil, lubricating oil, wax, asphalt manufacture, or cracking to make more gasoline and fuel oil.

The modern technical evaluation of crude oil by a purchaser depends upon many factors which are specific to each refinery A complete evaluation requires much information on the composition and properties of the fractions of the crude obtained on distillation and a knowledge of the probable corrosive effects of the crude

Whereas 20 years ago crude oils were bought and sold primarily on the basis of specific gravity, percentage of sulfur, and distillation data, the modern trend is to examine selected fractions by special techniques such as mass spectrometry, infrared spectroscopy, ultraviolet spectroscopy, gas chromatography, or physical property correlations to determine the suitability for reforming, cracking, and lubricating oil production. Each large purchaser of crude oil has developed a particular system Although the methods used are far from uniform, the objective is in each case to obtain sufficient information about the composition of the crude oil to predict how the fractions will behave in modern refinery operations See PETROLEUM.

Analysis for composition. Hydrocarbon mixtures lighter than C_4 are analyzed for composition by low temperature fractional distillation or gas-liquid chromatography followed by mass spectroscopy on the lightest cut to distinguish H_2 and CH_4

Fractions from C_4 to C[illegible] are separated by gas-liquid chromatography which identifies most of the species. Additional information is obtained by spectroscopic examination of any ambiguous cuts separated by chromatography

Heavier fractions require much more elaborate procedures although the distribution of hydrocarbon types within a narrow range of molecular weight can be approximated by fractional distillation or chromatography, followed by spectroscopic analysis of the fractions

Olefins, aromatics, and saturates in gasoline and jet fuels can be separated by percolating through a tube of silica gel using fluorescent dyes to mark the boundaries This is the FIA analysis

The mass spectrometer is especially useful for the determination of naphthenes in saturated fractions, because naphthenes with two rings can be distinguished from those with one ring

In very heavy oils, little can be done to identify specific hydrocarbons Such information as the total saturates (separated by silica gel) and the concentration of various aromatic types (determined by spectroscopy) can be obtained

Carbon-type composition is so closely connected with the physical properties of oils that it can be derived from such properties by a variety of correlations which are accurate, over their range of usefulness, to about one-half of one carbon atom. The correlation based on density, refractivity intercept, and number of carbon atoms can be applied over the widest range of compositions.

Analysis for sulfur, nitrogen, and some other inorganic elements can be made on all fractions by well-known methods See PETROLEUM PROCESSING.

[S S K.]

Bibliography. American Society for Testing Materials (ASTM), *ASTM Standards on Petroleum Products and Lubricants*, 1959; ASTM, *Significance of ASTM Tests for Petroleum Products*, 3d ed 1956; ASTM [illegible] *Oils*, [illegible] ossini, [illegible] *from* [illegible], 1953; K. van Nes and H. A. Van Westen, *Aspects of the Constitution of Mineral Oil*, 1951; H I. Waterman, C. Boelhouwer, and J. Cornelissen, *Correlation Between Physical Constants and Chemical Structure*, 1958.

Oil and gas field development

The field development for petroleum. In the petroleum industry, a field means an area underlain without substantial interruption by one or more reservoirs of commercially valuable oil or gas, or both. A single reservoir (or group of reservoirs which cannot be separately produced) is a pool. Several pools separated from one another by barren, impermeable rock may be superimposed one above another within the same field. Pools have variable areal extent Any sufficiently deep well located within the field should produce from one or more pools. However, each well cannot produce from every pool, because different pools have different areal limits. Development of a field includes the location, drilling, completion, and equipment of wells necessary to produce the commercially recoverable oil and gas in the field.

Related oil field conditions. Petroleum is a generic term which, in its broadest meaning, includes all naturally occurring hydrocarbons, whether gaseous, liquid, or solid. By variation of the temperature, pressure, or both, of any hydrocarbon, it becomes gaseous, liquid, or solid. Temperatures in producing horizons vary from 60° to more than 300°F, depending chiefly upon the depth of the horizon. A rough approximation is that temperature in the reservoir sands, or pay, equals 60°F plus 0.017°F/ft of depth below surface. Pressure on the hydrocarbons varies from atmospheric to more than 11,000 psi. Normal pressure is considered as 0.465 psi/ft of depth. Tem-

peratures and pressure vary widely from these average figures. Hydrocarbons, because of wide variations in pressure and temperatures and because of mutual solubility in one another, do not necessarily exist underground in the same phases in which they appear at the surface.

Petroleum occurs underground in porous rocks of wide variety. Commonly the oil and gas industry refers to separate accumulations of petroleum as pools of oil or gas. Actually, the petroleum is dispersed in rock in pore spaces and small openings ranging from microscopic size to rare holes 1 in. or more in diameter. The containing rock is commonly called the sand or the pay, regardless of whether the pay is actually sandstone, limestone, dolomite, unconsolidated sand, or fracture openings in relatively impermeable rock.

At the wellhead, petroleum is commonly separated into gas, crude oil, and condensate. All gas contains some liquefiable hydrocarbons which may be removed as natural gasoline and liquefied products in natural-gasoline plants. In some high-pressure gas pools, important volumes of liquids are recovered at the wellhead as condensate. All crude oil contains gas. In most large oil fields, the volume of gas makes profitable the gathering, processing, and ultimate marketing of the gas associated with the oil Such gas is commonly referred to as casinghead gas or oil-well gas. About one-third of the United States marketed production of gas comes from casinghead gas. Two-thirds of the marketed gas comes from gas wells and gas condensate wells. Casinghead gas and gas-well gas, after being stripped of their natural gasoline and liquefiable gas products, become the natural gas of commerce *See* NATURAL GAS; PETROLEUM PROCESSING.

Development of field. After discovery of a field producing oil or gas or both in commercial quantities, the field must be explored to determine its vertical and horizontal limits and the mechanisms under which the field will produce Development and exploitation of the field proceed simultaneously. Usually, the original development program is repeatedly modified by knowledge acquired during the exploration and exploitation of the field.

Ideally, tests should be drilled to the lowest possible producing horizon in order to determine the number of pools existing in the field. Testing of the first wells sometimes indicates the producing mechanisms and so indicates the best development program. Very early in the history of the field, step-out wells will be drilled to determine the areal extent of the pool or pools. Step-out wells give further information regarding the volumes of oil and gas

United States petroleum production in 1958

Product	Volume	10^{12} Btu	% of Btu
Crude oil, bbl	2,372,730,000	13,524	50 4
Natural gas liquids, bbl	341,518,000	1,435	5 3
Natural gas, 10^3 ft³	11,485,026,000	11,887	44 3
Total		26,846	100 0

available, the producing mechanisms, and the desirable spacing of wells.

The operator of an oil and gas field endeavors to select a development program which will produce the largest volume of oil and gas at a profit. The program adopted is always a compromise between conflicting objectives The operator desires (1) to drill the fewest wells which will efficiently produce the recoverable oil and gas, (2) to drill, complete, and equip the wells at the lowest possible cost, (3) to complete production in the shortest practical time to reduce both capital and operating charges, (4) to operate the wells at the lowest possible cost, and (5) to recover the largest possible volume of oil and gas.

Selecting the number of wells Oil pools are produced by four mechanisms: dissolved gas expansion, gas-cap drive, water drive, and gravity drainage (*see* PETROLEUM RESERVOIR ENGINEERING) Commonly, two or more mechanisms operate in a single pool. The type of producing mechanism in each pool influences the decision as to the number of wells to be drilled. Theoretically, a single, perfectly located well in a water-drive pool is capable of producing all of the commercially recoverable oil and gas from that pool. Practically, more than one well is necessary if a pool of more than 8 acres is to be depleted in a reasonable time. If pool produces under either gas expansion or gas-cap drive, oil production from the pool will be independent of the number of wells up to a spacing of at least 80 acres per well (1866 ft between wells) Gas wells are spaced a mile or more apart The operator accordingly selects the widest spacing permitted by field conditions and legal requirements as discussed later.

Major components of cost Costs of drilling, completing, and equipping the wells influence development plans Having determined the number and depths of producing horizons and the producing mechanisms in each horizon, the operator must decide whether he will drill a well at each location to each horizon or whether a single well may produce from two or more horizons at the same location. Clearly, the cost of drilling the field can be sharply reduced if a well can drain two, three, or more horizons (pools). The cost of drilling a well will be higher if several horizons are simultaneously produced, because the dual or triple completion of a well usually requires a larger hole and larger casing than would be necessary if the well produced from a single horizon. Further, completion and operating costs are higher if a well produces simultaneously from two or more horizons However, the increased cost of drilling a well of larger diameter and completing the well in two or more horizons is 20–40% less than the cost of drilling and completing two wells to produce separately from two horizons

In some cases, the operator may reduce the number of wells by drilling a well to the lowest producible horizon and taking production from that level until the horizon (pool) is there commercial

exhausted The well is then plugged back to produce from a higher horizon and perforated opposite this second zone. Successively, different horizons are depleted until no oil or gas reservoir remains available to the well.

Selection of the plan for producing the various horizons obviously affects the cost of drilling and completing individual wells, as well as the number of wells which the operator will drill If two wells are drilled at approximately the same location, they are referred to as twins, three wells at the same location are triplets, and so on.

Costs and duration of production The operator wishes to produce as rapidly as possible because the present worth of the money received from sale of hydrocarbons is obviously reduced in proportion as the life of the well is extended The successful operator must recover from his productive wells the costs of drilling and operating those wells, and in addition he must recover all costs involved in geological and geophysical exploration, leasing, scouting, drilling of dry holes, and occasionally others (*see* GEOPHYSICAL EXPLORATION, PROSPECTING, PETROLEUM). If profits from production are not sufficient to recover all exploration and production costs and yield a profit in excess of the rate of interest which the operator could secure from a different type of investment, the operator is discouraged from further exploration Accordingly, the operator produces the oil and gas as rapidly as practicable

Most wells cannot operate at full capacity, because unlimited production results in physical waste and sharp reduction in ultimate recovery. In many areas, conservation restrictions are enforced to make certain that the operator does not produce in excess of the maximum efficient rate For example, if an oil well produces at its highest possible rate, a zone promptly develops around the well where production is occurring under gas-expansion drive, the most inefficient producing mechanism. Slower production may permit the petroleum to be produced under gas-cap drive or water drive, in which case ultimate production of oil will be two to four times as great as it would be under gas-expansion drive Excessive production rate may cause water coning, loss of oil or gas, and premature abandonment of a well Accordingly, the most rapid rate of production is not necessarily the most efficient rate The operator cannot incur a loss of more than 50% of his oil or gas and still secure maximum recovery or maximum profit from the well

Similarly, the initial exploration of the field may indicate that one or more gas-condensate pools exist, and recycling may be necessary to secure maximum recovery both of condensate and of gas The decision to recycle will affect the number of wells, the locations of the wells, and the completion methods adopted in the development program

Further, as soon as the operator determines that secondary oil recovery methods are desired and expects to inject water, gas, or, rarely, air in order to provide additional energy to flush or displace oil from the pay, the number and location of wells may be modified to permit the most effective secondary recovery procedures

Legal and practical restrictions. The preceding discussion has assumed control of an entire field under a single ownership by a single operator. In the United States, a single operator rarely controls a large field and this field is almost never under a single lease. Usually, the field is covered by separate leases owned and operated by different producers. The development program must then be modified in consideration of the lease boundaries and the practices of other operators that are in the field.

Oil and gas know no lease boundaries. They move freely underground from areas of high pressure toward lower pressure situations. The operator of a lease is obligated to locate his wells in such a way as to prevent drainage of his lease by wells on adjoining leases, even though he may own the adjoining leases. In the absence of conservation restrictions, an operator must produce petroleum from his wells as rapidly as it is produced from wells on adjoining leases. Slow production on one lease results in migration of oil and gas to nearby leases which are more rapidly produced.

The operator's development program must provide for offset wells located as close to the boundary of his lease as are wells on adjoining leases Further, the operator must equip his wells to produce as rapidly as the offset produces and must produce from the same horizons (pools) which are being produced in offset wells. The lessor who sold the lease to the operator is entitled to his share of the recoverable petroleum underlying his land Negligence by the operator in permitting drainage of a lease makes the operator liable to suit for damages or cancellation of the lease

A development ...

... upon the best development program for a field Most states have enacted statutes and have appointed regulatory bodies ...

... of secondary recovery methods.

Drilling unit. Commonly, small leases or portions of two or more leases are combined to form a drilling unit in the center of which a well will be drilled Unitization may be voluntary, by agreement between the operator or operators and the interested royalty owners, with provision for sharing production from the well between the parties in proportion to their acreage interests. In many states the regulatory body has authority to require unitization of drilling units which eliminates unnecessary offset wells and protects the interests of a landowner whose acreage holding may be too small to justify the drilling of a single well on his property alone

Pool unitization When recycling or some types of secondary recovery are planned, further unitization is adopted. Since oil and gas move freely across lease boundaries, it would be wasteful for an operator to repressure, recycle, or water-drive a lease if the adjoining leases were not similarly operated. Usually, an entire pool must be unitized for efficient recycling or secondary recovery operations. Pool unitization may be accomplished by agreement between operators and royalty owners In many cases, difference of opinion or ignorance on the part of some parties prevents voluntary pool unitization. Many states authorize the regulatory body to unitize a pool compulsorily on application by a specified percentage of interests of operators and royalty owners. Such compulsory unitization is planned to provide each operator and each royalty owner his fair share of the petroleum products produced from the field regardless of the location of the well or wells through which these products actually reach the surface.

For further discussion of problems involved in oil and gas field development, *see* OIL AND GAS FIELD EXPLOITATION; OIL AND GAS WELLS; PETROLEUM SECONDARY RECOVERY. [R.S.K.]

Bibliography: E L. DeGolyer (ed.), *Elements of the Petroleum Industry*, 1940; B. W. Murphy (ed), *Conservation of Oil and Gas, A Legal History*, 1949; M Muskat, *Physical Principles of Oil Production*, 1949; L C Uren, *Petroleum Production Engineering—Oil Field Development*, 4th ed., 1956.

Oil and gas field exploitation

The complex of field production methods for petroleum. Oil and gas production necessarily are intimately related since approximately one-third of the gross gas production in the United States is produced from wells that are classified as oil wells However, the naturally occurring hydrocarbons of petroleum are not only liquid and gaseous but may even be found in a solid state, such as asphaltite and some asphalts *See* ASPHALT AND ASPHALTITE.

GENERAL CONSIDERATIONS

Where gas is produced without oil, the production problems are simplified because the product flows naturally throughout the life of the well and does not have to be lifted to the surface However, there are sometimes problems of water accumulations in gas wells and it is necessary to pump the water from the wells to maintain maximum, or economical, gas production. The line of demarcation between oil wells and gas wells is not definitely established since oil wells may have gas/oil ratios ranging from a few cubic feet per barrel to many thousand cubic feet of gas per barrel of oil. Most gas wells produce quantities of condensable vapors, such as propane and butane, that may be liquefied and marketed for fuel, and the more stable liquids produced with gas can be utilized as natural gasoline. *See* PETROLEUM.

Factors of method selection. The method selected for recovering oil from a producing formation depends on many factors, including well depth, well-casing size, oil viscosity, density, water production, gas/oil ratio, porosity and permeability of the producing formation, formation pressure, water content of producing formation, and whether the force driving the oil into the well from the formation is primarily gas pressure, water pressure, or a combination of the two Other factors such as paraffin content and difficulty expected from paraffin deposits, sand production, and corrosivity of the well fluids, also have a decided influence on the most economical method of production.

Special techniques utilized to increase productivity of oil and gas wells include acidizing hydraulic fracturing of the formation, the setting of screens, and gravel or sand packing to increase permeability around the well bore.

Aspects of production rate. Productive rates per well may vary from a few barrels per day to several thousand barrels per day, and it may be necessary to produce a large percentage of water along with the oil

Field and reservoir conditions. In some cases reservoir conditions are such that some of the wells will flow naturally throughout the entire economical life of an oil field. However, in the great majority of cases it is necessary to resort to artificial lifting methods at some time during the life of the field, and often it is necessary to apply artificial lifting means immediately after the well is drilled

Market and regulatory factors. In some oil producing states of the United States there are state regulatory bodies authorized to regulate oil production from the various oil fields. The allowable production per well is based on various factors, including the market for the particular type of oil available, but very often the allowable production is based on an engineering study of the reservoir to determine the optimum rate of production, thereby assuring maximum utilization of reservoir energy and maximum ultimate recovery from the producing formation The selection of the most economical production equipment depends, therefore, upon a great number of factors, some of which are not directly related to the productive capacity of the formation

The total crude oil production in the United States in 1957 has been estimated by the Bureau of Mines at 2.616.778.000 barrels (bbl), total world production for 1957 was estimated at 6.446 029.500 bbl Natural gas production in the United States in 1957 is reported by the American Gas Association as follows: net production (equals gross production minus amount returned to formation) 11,554,800,000,000 ft^3, marketed production (equals net production minus losses and waste) 10,629,200,000,000 ft^3.

Useful terminology. A few definitions of terms used in petroleum production technology are listed

below to assist in an understanding of some of the problems involved.

Porosity. The percentage porosity is defined as the percentage volume of voids per unit total volume. This, of course, represents the total possible volume available for accumulation of fluids in a formation, but only a fraction of this volume may be effective for practical purposes because of possible discontinuities between the individual pores.

Permeability. Permeability is a measure of the resistance to flow through a porous medium under the influence of a pressure gradient. The unit of permeability commonly employed in petroleum production technology is the darcy. A porous structure has a permeability of 1 darcy if, for a fluid of 1 centipoise (cp) viscosity, the volume flow is 1 $cm^3/(sec)(cm^2)$ under a pressure gradient of 1 atm/cm.

Productivity index. The productivity index is a measure of the capacity of the reservoir to deliver oil to the well bore through the productive formation and any other obstacles that may exist around the well bore. In petroleum production technology, productivity index is defined as production in barrels per day per pound drop in bottom hole pressure. For example, if a well is closed in at the casing head, the bottom hole pressure will equal the formation pressure when equilibrium conditions are established. However, if fluid is removed from the well, either by flowing or pumping, the bottom hole pressure will drop as a result of the resistance to flow of fluid into the well from the formation to replace the fluid removed from the well. If the closed in bottom hole pressure should be 1000 psi, for example, and if this pressure should drop to 900 psi when producing at a rate of 100 bbl/day (a drop of 100 psi), the well in question would have a productivity index of one.

Barrel. The standard barrel used in the petroleum industry is 42 U.S. gal.

API gravity. The American Petroleum Institute (API) scale that is in common use for indicating specific gravity or a rough indication of quality of crude petroleum oils, differs slightly from the Baumé scale commonly used for other liquids lighter than water. The following table shows the relationship between degrees API and specific gravity referred to water at 60°F for specific gravities ranging from 0.60 to 1.0.

Viscosity range. Viscosity of crude oils currently produced varies from approximately 1 cp to values above 1000 cp at temperatures existing at the bottom of the well. In some areas it is necessary to supply heat artificially down the wells or circulate lighter oils to mix with the produced fluid for maintenance of a relatively low viscosity throughout the temperature range to which the product is subjected. Bottom-hole or formation temperatures range from approximately 80 to 300°F, and even higher in extremely deep wells.

In addition to wells that are classified as gas wells or oil wells, the term gas-condensate well has come into general use during the past several years to designate a well that produces large volumes of gas with appreciable quantities of light condensable hydrocarbon fluids. Some of these fluids are liquid at atmospheric pressure and temperature; others, such as propane and butane, are readily condensed under relatively low pressures in gas separators for use as liquid petroleum gas (LPG) fuels or for other uses. The liquid components of the production from gas-condensate wells generally arrive at the surface in the form of small droplets entrained in the high-velocity gas stream and are separated from the gas in a high-pressure gas separator.

PRODUCTION METHODS IN PRODUCING WELLS

The common methods of producing oil wells are (1) natural flow, (2) pumping with sucker rods to actuate a pump located in the well fluid; (3) gas lift, (4) hydraulic subsurface pumps, in which the subsurface unit consists of a reciprocating hydraulic motor connected [illegible] driven by [illegible] the well to [illegible] trically driven centrifugal well pumps; and (6) swabbing.

Numerous other methods, including jet pumps and sonic pumps, have been tried and are used to slight extent. The sonic pump is a development in which the tubing is vibrated longitudinally by a mechanism at the surface and acts as a high-speed pump with an extremely short stroke.

As of December 31, 1957, *World Oil* reported a total of 563,710 producing oil wells in the United States. Of this number, 68,221, or 12.1%, were flowing wells, and 495,489, or 87.9%, were on artificial lift.

A brief discussion of production methods, in the approximate order of their relative importance and popularity, follows.

Degrees API corresponding to specific gravities of crude oil at 60°/60°F

Specific gravity in tenths	Specific gravity, in hundredths									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.60	104.33	100.47	96.73	93.10	89.59	86.19	82.89	79.69	76.59	73.57
0.70	70.64	67.80	65.03	62.34	59.72	57.17	54.68	52.27	49.91	47.61
0.80	45.38	43.19	44.06	38.98	36.95	34.97	33.03	31.14	29.30	27.49
0.90	25.72	23.99	22.30	20.65	19.03	17.45	15.90	14.38	12.89	11.43
1.00	10.00									

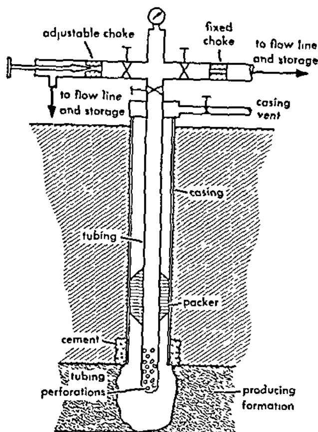


Fig. 1. Schematic view of well equipped for producing by natural flow.

Natural flow. Natural flow is the most economical method of production and generally is utilized as long as the desired production rate can be maintained by this method. It utilizes the formation energy, which may consist of gas in solution in the oil in the formation; free gas under pressure acting against the liquid and gas-liquid phase to force it toward the well bore; water pressure acting against the oil; or a combination of these three energy sources. In some areas the casing-head pressure may be of the order of 10,000 psi, so it is necessary to provide fittings adequate to withstand such pressures. Adjustable throttle valves, or chokes, are utilized to regulate the flow rate to a desired and safe value. With such a high pressure drop across a throttle valve the life of the valve is likely to be very short. Several such valves are arranged in parallel in the tubing head "Christmas tree" with positive shutoff valves between the chokes and the tubing head so that the wearing parts of the throttle valve, or the entire valve, can be replaced while flow continues through another similar valve.

An additional safeguard that is often used in connection with high-pressure flowing wells is a bottom-hole choke or a bottom-hole flow control valve that limits the rate of flow to a reasonable value, or stops it completely, in case of failure of surface controls. Figure 1 shows a schematic outline of a simple flowing well hook-up. The packer is not essential but is often used to reduce the free gas volume in the casing

Flow rates for domestic wells seldom exceed a few hundred barrels per day (bpd) because of either enforced or voluntary restrictions to regulate production rates and to obtain most efficient and economical ultimate recovery. However, in some foreign countries, especially in the Middle East, it is not uncommon for natural flow rates to exceed 10,000 bpd/well.

Lifting. Most wells are not self-flowing Eight types of lifting are outlined here

Pumping with sucker rods. Approximately 90% of the wells made to produce by some artificial lift method in the United States are equipped with sucker-rod-type pumps. In these the pump is installed at the lower end of the tubing string and is actuated by a string of sucker rods extending from the surface to the subsurface pump. The sucker rods are attached to a polished rod at the surface. The polished rod extends through a stuffing box and is attached to the pumping unit which produces the necessary reciprocating motion to actuate the sucker rods and the subsurface pump Figure 2 shows a simplified schematic section through a pumping well. The two common variations are (1) mechanical and (2) hydraulic long-stroke pumping.

1. Mechanical pumping. The great majority of pumping units are of the *mechanical* type, consisting of a suitable reduction gear, and crank and Pitman arrangement to drive a walking beam to produce the necessary reciprocating motion A counterbalance is provided to equalize the load on the upstroke and downstroke. Mechanical pumping units of this type vary in load-carrying capacity from about 2000 to about 43,000 lb and the torque rating of the low-speed gear which drives the crank

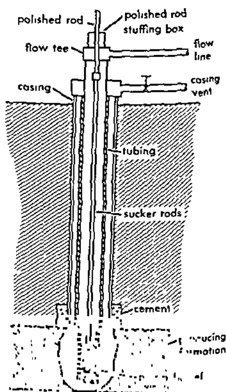


Fig. 2. Schematic view of well equipped for pumping with sucker rods

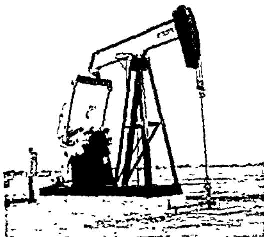


Fig 3 Modern pumping unit with adjustable rotary counterbalance (Oil Well Supply Division, U.S. Steel Corp)

ranges from 6400 in.-lb in the smallest API standard unit to about 1,500,000 in.-lb for the largest units now in use Stroke length varies from about 18 to 192 in Usual operating speeds are from about 6 to 20 strokes/min However, both lower and higher rates of speed are sometimes used Figure 3 shows a modern pumping unit in operation

Production rates with sucker-rod-type pumps vary from a fraction of 1 bpd in some areas, with part time pumping, to approximately 3000 bpd for the largest installations in relatively shallow wells

(2) Hydraulic long-stroke pumping. For this the units consist of an hydraulic lifting cylinder mounted directly over the well head and are designed to produce stroke lengths of as much as 30 ft Such long-stroke hydraulic units are usually equipped with a pneumatic counterbalance arrangement which equalizes the power requirement on the upstroke and downstroke.

Hydraulic pumping units also are made without any provision for counterbalance However, these units are generally limited to relatively small wells and they are relatively inefficient

Gas lift Gas lift in its simplest form consists of initiating or stimulating well flow by injecting gas at some point below the fluid level in the well With large-volume gas-lift operations the well may be produced through either the casing or the tubing In the former case gas is conducted through the tubing to the point of injection, in the latter, gas may be conducted to the point of injection through the casing or through an auxiliary string of tubing When gas is injected into the oil column, the weight of the column above the point of injection is reduced as a result of the space occupied by the relatively low density of the gas and this lightening of the fluid column is sufficient to permit the formation pressure to initiate flow up the tubing to the surface Gas injection is often utilized to increase the flow from wells that will flow naturally but will not produce the desired amount by natural flow.

There are many factors determining the advisability of adopting gas lift as a means of production. One of the more important factors is the availability of an adequate supply of gas at suitable pressure and reasonable cost. In a majority of cases gas lift cannot be used economically to produce a reservoir to depletion because the well may be relatively productive with a low back pressure maintained on the formation but will produce very little, if anything, with the back pressure required for gas-lift operation. Therefore, it generally is necessary to resort to some mechanical means of pumping before the well is abandoned and it may be more economical to adopt the mechanical means initially than to install the gas-lift system while conditions are favorable and later replace it.

This discussion of gas lift has dealt primarily with the simple injection of gas, which may be continuous or intermittent There are numerous modifications of gas-lift installations, including various designs of flow valve which may be installed in the tubing string to open and admit gas to the tubing from the casing at a predetermined pressure differential between the tubing and casing. When the valve opens, gas is injected into the tubing to initiate and maintain flow until the tubing pressure drops to a predetermined value and the valve closes before the input gas/oil ratio becomes excessive This represents an intermittent-flow type of valve. Other types are designed to maintain continuous flow, proper pressure differential, and proper gas injection rate for efficient operation. In some cases several such flow valves are spaced up the tubing string to permit flow to be initiated from various levels as required

Other modifications of gas lift involve the utilization of displacement chambers. These are installed on the lower end of the well tubing in which oil may accumulate, and the oil is displaced up the tubing with gas injection controlled by automatic or mechanical valves.

Hydraulic subsurface pumps The hydraulic subsurface pump has come into fairly prominent use in recent years, the subsurface pump is operated by means of a hydraulic reciprocating motor attached to the pump and installed in the well as a single unit. The hydraulic motor is driven by a supply of hydraulic fluid under pressure that is circulated down a string of tubing and through the motor. Generally the hydraulic fluid consists of crude oil which is discharged into the return line and returns to the surface along with the produced crude oil.

Hydraulically operated subsurface pumps are also arranged for separating the hydraulic power fluid from the produced well fluid This arrangement is especially desirable where the fluid being produced is corrosive or is contaminated with considerable quantities of sand or other solids that are difficult to separate to condition the fluid for use as satisfactory power oil.

Centrifugal well pumps Electrically driven centrifugal pumps have been used to some extent, es-

pecially in large-volume wells of shallow or moderate depths. Both the pump and the motor are restricted in diameter to run down the well casing, leaving sufficient clearance for the flow of fluid around the pump housing. With the restricted diameter of the impellers the discharge head necessary for pumping a relatively deep well can be obtained only by using a large number of stages and operating at a relatively high speed. The usual rotating speed for such units is 3600 rpm and it is not uncommon for such units to have 50 or more pump stages The direct-connected electric motor must be provided with a suitable seal to prevent well fluid from entering the motor housing, and electrical leads must be run down the well casing to supply power to the motor.

Swabs. Swabs have been used for lifting oil almost since the beginning of the petroleum industry. They usually consist of a steel tubular body equipped with a check valve which permits oil to flow through the tube as it is lowered down the well with a wire line. The exterior of the steel body is generally fitted with flexible cup-type soft packing that will fall freely but will expand and form a seal with the tubing when pulled upward with a head of fluid above the swab Swabs are run into the well on a wire line to a point considerably below the fluid level and then lifted back to the surface to deliver the volume of oil above the swab. They are often used for determining the productivity of a well that will not flow naturally and for assisting in cleaning paraffin from well tubing. In some cases swabs are used to stimulate wells to flow by lifting, from the upper portion of the tubing, the relatively dead oil from which most of the gas has separated.

Bailers Bailers are used to remove fluids from wells and for cleaning out solid material They are run into the wells on wire lines as in swabbing but differ from swabs in that they generally are run only in the casing when there is no tubing in the well, and the capacity of the bailer itself represents the volume of fluid lifted each time since the bailer does not form a seal with the casing The bailer is simply a tubular vessel with a check valve in the bottom This check valve generally is arranged so that it is forced open when the bailer touches bottom in order to assist in picking up solid material for cleaning out a well

Jet pumps. A jet pump for use in oil wells operates on exactly the same principle as a water-well jet pump. Advantage is taken of the Bernoulli effect to reduce pressure by means of a high-velocity fluid jet. Thus oil is entrained from the well with this high-velocity jet in a venturi tube to accelerate the fluid and assist in lifting it to the surface, along with any assistance from the formation pressure. The application of jet pumps to oil wells has been insignificant to date.

Sonic pumps. Sonic pumps are a more recent development. Essentially they consist of a string of tubing equipped with a check valve at each joint and mechanical means on the surface to vibrate the tubing string longitudinally. This creates a harmonic condition that will result in several hundred strokes per minute with the strokes being a small fraction of 1 in. in length. Some of these pumps are now in use in relatively shallow wells, but it is perhaps too early to predict their ultimate field of application.

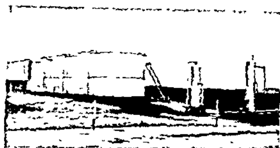


Fig. 4 Lease tank battery and gas separators. (Gulf Oil Corp.)

Lease tanks and gas separators. Figure 4 shows a typical lease tank battery consisting of four 1000-bbl tanks and two gas separators. Such equipment is used for handling production from wells produced by natural flow, gas lift, or pumping. In some pumping wells the gas content may be too low to justify the cost of separators for saving the gas.

Natural gasoline production. An important phase of oil and gas production in many areas is the production of natural gasoline from gas taken from the casing head of oil wells or separated from the oil and conducted to the natural gasoline plant The plant consists of facilities for compressing and extracting the liquid components from the gas The natural gasoline generally is collected by cooling and condensing the vapors after compression or by absorbing in organic liquids having high boiling points from which the volatile liquids are distilled Many natural gasoline plants utilize a combination of condensing and absorbing techniques. Figure 5 shows an over-all view of a modern natural gasoline plant operating in western Texas.

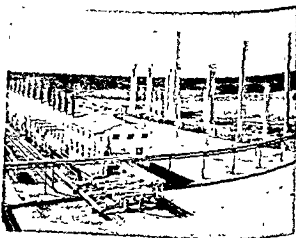


Fig. 5 Modern natural gasoline plant in western Texas. (Gulf Oil Corp.)

PRODUCTION PROBLEMS AND INSTRUMENTS

Six major problems and six kinds of instruments are outstanding in field exploitation

Corrosion. In many areas the corrosion of production equipment is a major factor in the cost of petroleum production The following comments on the oil field corrosion problem are taken largely from, and reproduced by permission from, NACE-API *Corrosion of Oil and Gas-Well Equipment*

For practical consideration, corrosion in oil- and gas-well production can be classified into four main types

1 Sweet corrosion occurs as a result of the presence of carbon dioxide and fatty acids. Oxygen and hydrogen sulfide are not present This type of corrosion occurs in both gas-condensate and oil wells It is most frequently encountered domestically in southern Louisiana and Texas, and other scattered areas. At least 20% of all sweet oil production and 45% of condensate production is considered corrosive

2. Sour corrosion is designated as corrosion in oil and gas wells producing even trace quantities [illegible]

Arbuckle production in Kansas and in the Permian Basin of western Texas and New Mexico About 12% of all sour production is considered corrosive.

3 Oxygen corrosion occurs wherever equipment is exposed to atmospheric oxygen It occurs most frequently in offshore installations, brine-handling and injection systems, and in shallow producing wells where air is allowed to enter the casing

4 Electrochemical corrosion is designated as that which occurs when corrosion currents can be readily measured or when corrosion can be mitigated by the application of current, as in soil corrosion

Corrosion inhibitors are used extensively in both oil and gas wells to reduce corrosion damage to subsurface equipment Most of the inhibitors used in the oil field are of the so-called polar organic type All of the major inhibitor suppliers can now furnish effective inhibitors for the prevention of sweet corrosion as encountered in most fields These can be purchased in oil-soluble, water-dispersible or water-soluble form

Paraffin deposits. In many crude-oil-producing areas paraffin [illegible] on sucker [illegible] the and ex[illegible] the tubing [illegible] or completely plugged. It is necessary to remove these deposits to maintain production rates A variety of methods are used to remove paraffin from the tubing including the application of heated oil through tubular sucker rods to mix with and transfer heat to the oil being produced and raise the temperature to a point where the deposited paraffin will be dissolved or melted Paraffin solvents may also be applied in this manner without the necessity of applying heat. Mechanical means, often are used in which a scraping tool is run on a wire line and paraffin is scraped from the tubing wall as the tool is pulled back to the surface. Mechanical scrapers that attach to sucker rods also are in common use. Various types of automatic scraper have been used in connection with flowing wells. These consist of a form of piston that will drop freely to the bottom when flow is stopped but will rise back to the surface when flow is resumed Electrical heating methods have been used rather extensively in some areas The tubing is insulated from the casing and from the flow line, and electric current is transmitted through the tubing for the time necessary to heat the tubing sufficiently to cause the paraffin deposits to melt or go into solution in the oil in the tubing Plastic coatings have been utilized inside tubing and flow lines to minimize or prevent paraffin deposits Paraffin does not deposit readily on certain plastic coatings.

A common method for removing paraffin from flow lines is to disconnect the line at the well head and at the tank battery and force live steam through the line to melt the paraffin deposits and flow them out. Various designs of flow-line scrapers have also been used rather extensively and fairly successfully Paraffin deposits in flow lines are minimized by insulating the lines or by burying the lines to maintain a higher average temperature.

Emulsions. A large percentage of oil wells produce various quantities of salt water along with the oil, and numerous wells are being pumped where the salt-water production is 90% or more of the total fluid [illegible] tion met[illegible] of water [illegible] The more common type is oil-in water Emulsions are treated with a variety of demulsifying chemicals, with the application of heat, and a combination of these two treatments Another method for breaking emulsions is the electrostatic or electrical precipitator type of emulsion treatment In this method the emulsion to be broken is circulated between electrodes subjected to a high potential difference The resulting concentrated electric field tends to rupture the oil-water interface and thus breaks the emulsion and permits the water to settle out. Figure 6 shows two pumping wells with a tank battery in the background This tank battery is equipped with a wash tank, or gun barrel, and a gas-fired heater for emulsion treating and water separation before the oil is admitted to the lease tanks.

Gas conservation. If the quantity of gas produced with crude oil is appreciably greater than can be efficiently utilized or marketed, it is necessary to provide facilities for returning the excess gas to the producing formation. Formerly, large quantities of excess gas were disposed of by burning or simply by venting to the atmosphere. This practice is now unlawful. Returning excess gas to the formation not only conserves the gas for future use but also results in greater ultimate recovery of oil from the formation

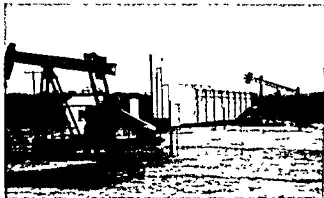


Fig 6. Two pumping wells with tank battery. Tank battery is equipped with gun barrel and heater for treating emulsions. (*Oil Well Supply Division, U S. Steel Corp*)

Salt-water disposal. The large volumes of salt water produced with the oil in some areas present serious disposal problems It is generally pumped back to the formation through wells drilled for this purpose. Such salt-water disposal wells are located in areas where the formation already contains water and this practice helps to maintain the formation pressure as well as the productivity of the producing wells.

Offshore production. Offshore wells present additional production problems since the wells must be serviced from barges or boats. Wells of reasonable depth on land locations are seldom equipped with derricks for servicing because it is more economical to set up a portable mast for pulling and installing rods, tubing, and other equipment. However, the use of portable masts is not practical on offshore locations and a derrick is generally left standing over such wells throughout their productive life to facilitate servicing. There are a considerable number of offshore wells along the Gulf Coast and the Pacific Coast of the United States, but by far the greatest number of offshore wells in a particular region is in Lake Maracaibo in Venezuela. Figure 7 shows a considerable number of derricks in Lake Maracaibo with pumping wells in the foreground. These wells are pumped by electric power through cables laid on the lake bottom to conduct electricity from power-generating stations on shore. An over-water tank battery is visible at the extreme right. All offshore installations such as tank batteries, pump stations, and the derricks and pumping equipment, are supported on pilings in water up to 100 ft or more in depth. Currently there are approximately 2300 oil derricks in Lake Maracaibo. A growing number of semipermanent platform rigs and even bottom storage facilities are being used in Gulf of Mexico waters at depths up to more than 100 ft. *See* MINERAL FUEL AREAS; PETROLEUM GEOLOGY.

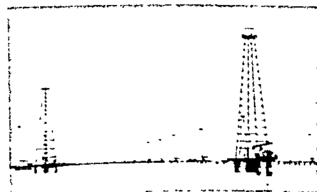


Fig 7. Offshore wells in Lake Maracaibo, Venezuela. (*Creole Petroleum Corp.*)

Instruments. The more common and more important instruments required in petroleum production operations are included among the following:

1. Gas meters, which are generally of the orifice type and are designed to record the differential pressure across the orifice, and the static pressure.

2. Recording subsurface pressure gages small enough to run down 2-in. ID (inside diameter) tubing are used extensively for measuring pressure gradients down the tubing of flowing wells, recording pressure build-up when the well is closed in, and measuring equilibrium bottom-hole pressures.

3. Subsurface samplers designed to sample well fluids at various levels in the tubing to determine physical properties such as viscosity, gas content, free gas, and dissolved gas at various levels. These instruments may also include a recording thermometer or a maximum reading thermometer, depending upon the information required

4. Oil meters of various types are utilized to meter crude oil flowing to or from storage.

5 Dynamometers for measuring polished rod loads. These instruments are sometimes known as well weighers since they are used to record the polished-rod load throughout a pumping cycle of a sucker-rod-type pump. They are used for determining maximum load on polished rods, as well as load variations, to permit accurate counterbalancing of pumping wells and to assure that pumping units or sucker-rod strings are not seriously overloaded

6. Liquid-level gages and controllers similar to those used in other industries for the same purpose but with special designs for closed lease tanks.

A wide variety of scientific instruments find application in petroleum production problems. The above outline gives an indication of a few specialized instruments used in this branch of the industry, and there are many more Special instruments developed by service companies are valued for a wide variety of purposes and include calipers to detect and measure corrosion pits inside tubing and casing, magnetic instruments to detect microscopic cracks in sucker rods, and other specialized instruments *See* OIL AND GAS STORAGE; PETROLEUM SECONDARY RECOVERY [R L CH]

Bibliography: W. F. Cloud, *Petroleum Production*, 1939; T. C. Frick, *Petroleum Production Handbook*, 1961; V. Guthrie, *Petroleum Products Handbook*, 1960; Nat. Assoc. Corrosion Eng and Am. Petroleum Inst., *Corrosion of Oil- and Gas well Equipment*, 1958; L. C. Uren, *Petroleum Production Engineering*, vol. 2, 3d ed., 1953

d gas storage

u of Mines, crude oil in storage during the of September 12, 1959, was 249,155,000 bbl duction, for the same week, processed-crude-storage in the form of stocks of gasoline, kero-distillate, and residual amounted to 441,688,-bbl

neral crude-oil storage. Crude-oil storage ties will be placed at a producing well for orary storage of individual well production, lease for gathering and storing all local oil uction until it can be treated or made avail-for shipping, at pipeline, pumping station, or inals, and at refinery locations

rude oil is stored in tanks or reservoirs The rican Petroleum Institute has adopted stand-

ucts, generally supplied with a low-pitched ical roof The smaller sizes of these tanks, gen-lly in pairs, take the oil directly from the pro-ing well and separator assembly and provide porary storage The larger tanks, in groups own as *batteries*, are used to store the entire tput on a lease, prior to delivery into a pipe-e transportation system Assemblages of the ger sizes of these steel tanks, known as tank rms, may be used for more permanent storage terminals or refinery locations

For the storage of very large amounts of crude l, particularly in tank farms reservoirs of con-rete construction may be used. These have been eveloped and used chiefly in California One such eservoir, elliptical in form, is 780 ft long 467 ft ide, and 23 ft deep It covers 9¼ acres of ground nd provides storage for more than 1 000 000 bbl f crude petroleum

Problems of crude-oil volatility. The principal technical factor involved in the storage of crude oil is its volatility The more volatile the oil, the higher will be the loss during storage caused by vapori-zation and the greater will be the pressure which the storage container must sustain Loss caused by vaporization is a matter of considerable economic importance and various devices are in use to reduce vaporization to a minimum Cylindrical tanks may have water tight, flat deck roofs a few inches below the top of the tank. The deck is covered with water in order to reduce the temperature variation within the tank, and in turn to reduce the vaporization of oil

Other cylindrical tanks are provided with roofs that float directly on the crude-oil surface, thus eliminating space within which vapor can ac-cumulate. Floating roofs require special devices for removing rainwater, for providing buoyancy, or for holding water on the rooftop to keep tem-perature variation to a minimum In some cases, the floating roof is provided with pontoons.

Still another type of tank has a flexible dia-phragm type roof capable of expanding and con-tracting as the vapor pressure within the tank varies The roofs are provided with safety control valves to prevent excess pressure build-up. Such a tank of 80,000-bbl capacity and 117-ft diameter may have a breathing roof rise of as much as

accomplished only by construction of closed tanks capable of withstanding vapor pressure of the stored fluid. Such tanks may need to withstand pressures of the order of 100 lb/in². These tanks, designed particularly for condensate fluids, natural gasoline hydrocarbons such as propane or butane, are generally spherical or spheroidal in form. High-pressure storage vessels are generally limited in capacity because of the amount of steel required

Underground storage. The storage of large amounts of natural gasoline or other highly volatile hydrocarbons is best achieved by underground stor-age Actual fluid in underground storage at the end

states, and when current projects underway are completed, total capacity will be in excess of 49,-000,000 bbl. Texas alone provides over 31,000,000 bbl of this capacity

Hydrocarbon liquids. Underground storage of hydrocarbon liquids may be in salt domes and salt layers; in mined caverns, consisting of shale, gran-ite, or limestone, in depleted oil and gas sands; or in water sands One salt-dome storage facility in Texas is reported to have a capacity of 3,600,000 bbl Although capacities of such facilities may be as low as 10,000 bbl, the majority have capacities of 100,000–500,000 bbl. Underground storage in salt layers may be at considerable depths, such as the Spearfish and Charles salt formations in North Dakota at 7000–8500 ft. Approximately 80% of the underground storage capacity for volatile hydrocarbon liquids is in salt domes and salt layers

Salt formation storage space is prepared by drilling wells to the salt formation with subsequent circulation of water to produce a cavity. Limestone, shale, or granite caverns are produced by standard mining methods In oil, gas, and water sands, the native porosity permits the storage of natural gas-oline materials.

Patterns of gas storage Natural gas is stored underground either to provide large and immedi-ately available amounts of gas for entry to high-pressure cross-country pipelines, or to provide large storage capacities near the point of con-

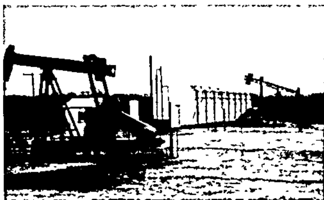


Fig. 6. Two pumping wells with tank battery. Tank battery is equipped with gun barrel and heater for treating emulsions. (*Oil Well Supply Division, U.S. Steel Corp*)

Salt-water disposal. The large volumes of salt water produced with the oil in some areas present serious disposal problems. It is generally pumped back to the formation through wells drilled for this purpose. Such salt-water disposal wells are located in areas where the formation already contains water and this practice helps to maintain the formation pressure as well as the productivity of the producing wells.

Offshore production. Offshore wells present additional production problems since the wells must be serviced from barges or boats. Wells of reasonable depth on land locations are seldom equipped with derricks for servicing because it is more economical to set up a portable mast for pulling and installing rods, tubing, and other equipment. However, the use of portable masts is not practical on offshore locations and a derrick is generally left standing over such wells throughout their productive life to facilitate servicing. There are a considerable number of offshore wells along the Gulf Coast and the Pacific Coast of the United States, but by far the greatest number of offshore wells in a particular region is in Lake Maracaibo in Venezuela. Figure 7 shows a considerable number of derricks in Lake Maracaibo with pumping wells in the foreground. These wells are pumped by electric power through cables laid on the lake bottom to conduct electricity from power-generating stations on shore. An over-water tank battery is visible at the extreme right. All offshore installations such as tank batteries, pump stations, and the derricks and pumping equipment, are supported on pilings in water up to 100 ft or more in depth. Currently there are approximately 2300 oil derricks in Lake Maracaibo. A growing number of semipermanent platform rigs and even bottom storage facilities are being used in Gulf of Mexico waters at depths up to more than 100 ft. *See* MINERAL FUEL AREAS; PETROLEUM GEOLOGY.

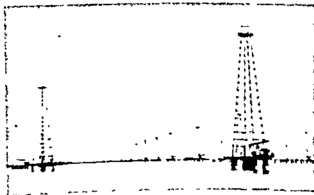


Fig. 7. Offshore wells in Lake Maracaibo, Venezuela. (*Creole Petroleum Corp.*)

Instruments. The more common and more important instruments required in petroleum production operations are included among the following.

1. Gas meters, which are generally of the orifice type and are designed to record the differential pressure across the orifice, and the static pressure.

2. Recording subsurface pressure gages small enough to run down 2-in. ID (inside diameter) tubing are used extensively for measuring pressure gradients down the tubing of flowing wells, recording pressure build-up when the well is closed in, and measuring equilibrium bottom-hole pressures.

3. Subsurface samplers designed to sample well fluids at various levels in the tubing to determine physical properties such as viscosity, gas content, free gas, and dissolved gas at various levels. These instruments may also include a recording thermometer or a maximum reading thermometer, depending upon the information required.

4. Oil meters of various types are utilized to meter crude oil flowing to or from storage.

5. Dynamometers for measuring polished rod loads. These instruments are sometimes known as well weighers since they are used to record the polished-rod load throughout a pumping cycle of a sucker-rod-type pump. They are used for determining maximum load on polished rods, as well as load variations, to permit accurate counterbalancing of pumping wells and to assure that pumping units or sucker-rod strings are not seriously overloaded.

6. Liquid-level gages and controllers similar to those used in other industries for the same purpose but with special designs for closed lease tanks.

A wide variety of scientific instruments find application in petroleum production problems. The above outline gives an indication of a few specialized instruments used in this branch of the industry, and there are many more. Special instruments developed by service companies are valued for a wide variety of purposes and include calipers to detect and measure corrosion pits inside tubing and casing, magnetic instruments to detect microscopic cracks in sucker rods, and other specialized instruments. *See* OIL AND GAS STORAGE; PETROLEUM SECONDARY RECOVERY. [R.L.CH.]

Bibliography: W. F. Cloud, *Petroleum Production*, 1939; T. C. Frick, *Petroleum Production Handbook*, 1961; V. Guthrie, *Petroleum Products Handbook*, 1960; Nat. Assoc. Corrosion Eng. and Am. Petroleum Inst., *Corrosion of Oil- and Gas-well Equipment*, 1958; L. C. Uren, *Petroleum Production Engineering*, vol. 2, 3d ed., 1953

Oil and gas storage

Storage, in quantities that may sometimes be very large, of crude oil and natural gas produced from their natural reservoirs. Storage is essential in order to have a produced reserve for emergency use, to meet seasonal or other fluctuations in demand for these raw materials, and to provide for efficient operation of the producing equipment, pipelines, and refineries. According to statistics of the U.S. Bureau of Mines, crude oil in storage during the week of September 12, 1959, was 249,155,000 bbl. In addition, for the same week, processed-crude-oil storage in the form of stocks of gasoline, kerosine, distillate, and residual amounted to 441,688,000 bbl.

General crude-oil storage. Crude-oil storage facilities will be placed at a producing well for temporary storage of individual well production, on a lease for gathering and storing all local oil production until it can be treated or made available for shipping, at pipeline, pumping station, or terminals, and at refinery locations.

Crude oil is stored in tanks or reservoirs. The American Petroleum Institute has adopted standards for the construction of both welded and riveted steel storage tanks [illegible]

[illegible] temporary storage. The larger tanks, in groups known as batteries, are used to store the entire output on a lease, prior to delivery into a pipeline transportation system. Assemblages of the larger sizes of these steel tanks, known as tank farms, may be used for more permanent storage at terminals or refinery locations.

For the storage of very large amounts of crude oil, particularly in tank farms, reservoirs of concrete construction may be used. These have been developed and used chiefly in California. One such reservoir, elliptical in form, is 780 ft long, 467 ft wide, and 23 ft deep. It covers 9⅔ acres of ground, and provides storage for more than 1,000,000 bbl of crude petroleum.

Problems of crude-oil volatility. The principal technical factor involved in the storage of crude oil is its volatility. The more volatile the oil, the higher will be the loss during storage caused by vaporization and the greater will be the pressure which the storage container must sustain. Loss caused by vaporization is a matter of considerable economic importance and various devices are in use to reduce vaporization to a minimum. Cylindrical tanks may have water tight, flat-deck roofs a few inches below the top of the tank. The deck is covered with water in order to reduce the temperature variation within the tank, and in turn to reduce the vaporization of oil.

Other cylindrical tanks are provided with roofs that float directly on the crude-oil surface, thus eliminating space within which vapor can accumulate. Floating roofs require special devices for removing rainwater, for providing buoyancy, or for holding water on the rooftop to keep temperature variation to a minimum. In some cases, the floating roof is provided with pontoons.

Still another type of tank has a flexible diaphragm type roof capable of expanding and contracting as the vapor pressure within the tank varies. The roofs are provided with safety control valves to prevent excess pressure build up. Such a tank of 80,000-bbl capacity and 117 ft diameter may have a breathing roof rise of as much as 24 in. above its supports, with an increase in vapor storage space of 12,000 ft³.

The complete elimination of evaporation loss is accomplished only by construction of closed tanks capable of withstanding vapor pressure of the stored fluid. Such tanks may need to withstand pressures of the order of 100 lb/in². These tanks, designed particularly for condensate fluids, natural gasoline hydrocarbons such as propane or butane, are generally spherical or spheroidal in form. High-pressure storage vessels are generally limited in capacity because of the amount of steel required.

Underground storage. The storage of large amounts of natural gasoline or other highly volatile hydrocarbons is best achieved by underground storage. Actual fluid in underground storage at the end of the summer of 1959 was reported by the *Oil and Gas Journal* to be 24,400,000 bbl. The same source reports total storage capacity of this type in 23 states, and when current projects underway are completed, total capacity will be in excess of 49,000,000 bbl. Texas alone provides over 31,000,000 bbl of this capacity.

Hydrocarbon liquids. Underground storage of hydrocarbon liquids may be in salt domes and salt layers, in mined caverns, consisting of shale, granite, or limestone; in depleted oil and gas sands; or in water sands. One salt-dome storage facility in Texas is reported to have a capacity of 3,600,000 bbl. Although capacities of such facilities may be as low as 10,000 bbl, the majority have capacities of 100,000–500,000 bbl. Underground storage in salt layers may be at considerable depths, such as the Spearfish and Charles salt formations in North Dakota at 7000–8500 ft. Approximately 80% of the underground storage capacity for volatile hydrocarbon liquids is in salt domes and salt layers.

Salt-formation storage space is prepared by drilling wells to the salt formation with subsequent circulation of water to produce a cavity. Limestone, shale, or granite caverns are produced by standard mining methods. In oil, gas, and water sands, the native porosity permits the storage of natural gasoline materials.

Patterns of gas storage. Natural gas is stored underground either to provide large and immediately available amounts of gas for entry to high-pressure cross-country pipelines, or to provide large storage capacities near the point of con-

sumption so that seasonal and daily demand fluctuations can be met. In the large metropolitan areas of the Northeast and Midwest, the consumption of natural gas is low during the summer months, but high on cold days in winter. Supplying natural gas to industry and to a large number of householders is not possible unless large storage capacity near the point of consumption can be procured, because pipelines cannot economically deliver gas over such wide demand ranges. The Committee on Underground Storage of the American Gas Association reports as of December 31, 1955, a total underground storage capacity for natural gas of 2,095,814,139,000 ft³.

Natural gas may be stored in what was originally a dry gas reservoir, an oil and gas reservoir, an oil reservoir, or an aquifer. The early underground natural gas storage was in the dry gas reservoirs of the Appalachian area These reservoirs at depths of 7000–8000 ft, in highly permeable, closed reservoirs, particularly of the Oriskany sand, were situated close enough to metropolitan markets to be attractive as storage locations. Of the 171 underground storage reservoirs active in 1955, 108 were in the Appalachian area, and of these, 101 were originally dry gas reservoirs.

Balancing reservoir storage. In a storage reservoir sufficient cushion gas must be maintained to provide pressure for well deliverability at desired rates. Between 40 and 50% of the total gas storage is considered to be cushion gas. The remainder, 50–60%, is active working gas. During the summer months, gas from a pipeline delivery source is injected into the reservoir, and the pressure is built up to its maximum working level sometime after the close of summer and before the advent of winter As gas consumption increases during the fall and winter months, gas is taken from storage The amount of gas in the storage reservoir declines until it reaches a minimum value sometime during the spring or early summer months Maximum pressure for storage of gas generally does not exceed the original indigenous reservoir pressure.

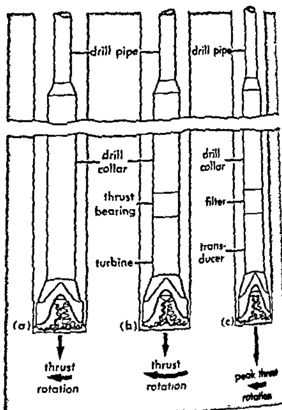
See OIL AND GAS FIELD EXPLOITATION; PETROLEUM ENGINEERING. [J.C.C.]

Bibliography: R. L. Huntington, *Natural Gas and Natural Gasoline*, 1950; D. L. Katz et al., *Handbook of Natural Gas Engineering*, 1959, L. C. Uren, *Petroleum Production Engineering: Exploitation*, 3d ed., 1953.

Oil and gas well drilling

Certain features are common to all oil- and gas-well drilling. Rock and sand formations of the earth must be penetrated, and the cuttings must be removed. The hole, or at least part of it, must then be cased, or lined, with pipe to prevent caving, to seal out water, and to permit production of oil or gas.

Two principal drilling methods use a rotary tool or a cable tool. The turbodrill method, which is a type of rotary drilling, is widely used in the Soviet ... accounted for nearly



Principles of (a) rotary drill, (b) turbodrill, and (c) vibratory drill. Arrows at bottom indicate relative importance of thrust and rotation.

90% of the 53,668 wells drilled. Because of it operating advantages, the older, cable tool metho is still dominant in Ohio, Pennsylvania, Kentuck West Virginia, and Michigan. Cable tools are us primarily in drilling shallow wells (usually abo 5000 ft) through hard formations, and frequentl to complete wells where low pressures are met i the subsurface formations

Hole diameter is related to depth, and decrea as the hole goes deeper For instance, a hole goi down 20,000 ft, may be as wide as 24 in. for t first few hundred feet and may taper to less th 8⅝ in at the bottom

In addition to the methods mentioned, drillin techniques involving variations in the manner conveying power to the bit, or cutting tool, are al under study These experimental techniques i clude vibration drilling, sonic drilling, and rotar percussion drilling. *See* CABLE-TOOL DRILL; R TARY TOOL DRILL; TURBODRILL. [A.L.P

Bibliography: L C Uren, *Petroleum Producti Engineering Oil Field Development*, 4th ed., 19

Oil and gas wells

Holes that have been formed mechanically in t earth's crust, penetrating a commercial subterr nean deposit or deposits of liquid or gaseous p troleum hydrocarbons or both, and so equipp that the formation product may be safely expell or withdrawn.

The first commercial oil well was drilled at T tusville, Pennsylvania, in ...9, to a total ... 69 ft. After 100 ...

ell was drilled to a depth of 25,340 ft. The first
ell sparked the commercial subterranean search
... 66.7% of the
... eum pro-
... segments;
... drilling
activities is referred to as the drilling department. The responsibility for successful drilling of the wells is delegated to this department, whether the actual drilling is done with company or contract *drilling equipment*. This group is composed of engineers and other trained people. The engineers provide the specifications for the wells and develop the program for the execution of the work. The drilling crew then takes over and carries out these programs. The drilling and direct supervision of a well require the services of approximately 16 men, divided into three 5-man crews, working 8-hour shifts around the clock, under the supervision of a drilling foreman or tool pusher. In addition to these men are those employed by the specialty service companies whose services are required for such operations as well logging, cementing, and perforating. *See* WELL LOGGING (MINERAL).

Boring and drilling. Two drilling techniques, churn and rotary drilling, are presently utilized. *See* BORING AND DRILLING, MINERAL.

Churn drilling. This method of drilling, more commonly known to the industry as cable tool drilling, utilizes an up-and-down motion of the drilling tools to fracture and crush the formation and mix it with water. This mixture is then dipped from the hole with a bailer which is run to the bottom of the hole on a wire line. Cable tool drilling is comparable in principle to a carpenter's making a hole in wood with a hammer and chisel. Although this system was first used on an oil well in 1859 and the system itself dates back to antiquity in drilling for salt, the method is not outmoded. Its use is generally restricted to shallow, low-pressure drilling.

Rotary drilling. Such drilling, which is comparable to a carpenter's drilling a hole in a piece of wood with a brace and bit or auger, was first introduced about 1900, and permitted a more economic and efficient means of drilling deeper wells. Rotary drilling utilizes the rotation of a drilling bit which is screwed ...
We ...
and ...
carried to the surface by drilling mud, which is circulated continuously during the drilling process. The mud is pumped into the drill pipe at the surface, out through the bit, and up the annular space between the drill pipe and the walls of the hole. At the surface, the drilled particles are removed from the mud by gravity settlement and shale shaker screens; the cleaned mud is again pumped into the hole in a continuous circulating process.

Power and power transmission. Three types of motive power are utilized: internal combustion engines, steam engines, and electric motors.

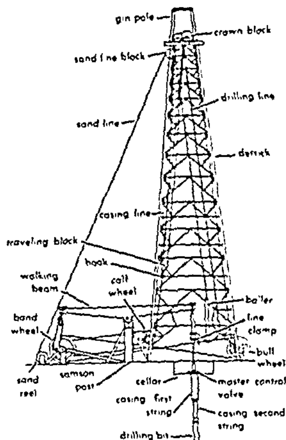


Fig. 1 Cable-tool rig.

In order to place a rotary rig in operation, a great deal of horsepower must be developed to rotate the drilling tool, lift and lower the drill pipe and casing, power the mud pumps, and supply light and power for other necessary equipment. This horsepower is usually developed at the drilling site by steam or internal combustion engines. The horsepower developed may be transmitted mechanically by chain or belt to the draw-works (hoist), mud pumps, rotary table, and other equipment, or may drive electrical generators which supply energy to electric motors which, in turn, drive the draw-works, pumps and other units.

The fuels used are gasoline, diesel oil, natural gas or liquefied petroleum gases such as propane and butane.

Circulating media. The purpose of a circulating medium is to cool and lubricate the bit, to remove the cuttings from the bore hole, and to protect the walls of the hole until the latter is cased. Usually the medium is a specially formulated fluid, compounded of several of many different ingredients, which vary with different areas and conditions, the basic ingredients of which are water and colloidal clay (bentonite), and is referred to in the industry as mud. *See* BENTONITE.

The physical character of the fluid must provide sufficient weight to control any gas or subsurface fluid pressures which may exist, and yet must be light enough to prevent mud loss to lower-pressure formations.

sumption so that seasonal and daily demand fluctuations can be met. In the large metropolitan areas of the Northeast and Midwest, the consumption of natural gas is low during the summer months, but high on cold days in winter. Supplying natural gas to industry and to a large number of householders is not possible unless large storage capacity near the point of consumption can be procured, because pipelines cannot economically deliver gas over such wide demand ranges. The Committee on Underground Storage of the American Gas Association reports as of December 31, 1955, a total underground storage capacity for natural gas of 2,095,-814,139,000 ft³.

Natural gas may be stored in what was originally a dry gas reservoir, an oil and gas reservoir, an oil reservoir, or an aquifer The early underground natural gas storage was in the dry gas reservoirs of the Appalachian area. These reservoirs at depths of 7000–8000 ft, in highly permeable, closed reservoirs, particularly of the Oriskany sand, were situated close enough to metropolitan markets to be attractive as storage locations Of the 171 underground storage reservoirs active in 1955, 108 were in the Appalachian area, and of these, 101 were originally dry gas reservoirs.

Balancing reservoir storage. In a storage reservoir sufficient cushion gas must be maintained to provide pressure for well deliverability at desired rates. Between 40 and 50% of the total gas storage is considered to be cushion gas. The remainder, 50–60%, is active working gas. During the summer months, gas from a pipeline delivery source is injected into the reservoir, and the pressure is built up to its maximum working level sometime after the close of summer and before the advent of winter. As gas consumption increases during the fall and winter months, gas is taken from storage The amount of gas in the storage reservoir declines until it reaches a minimum value sometime during the spring or early summer months Maximum pressure for storage of gas generally does not exceed the original indigenous reservoir pressure

See OIL AND GAS FIELD EXPLOITATION; PETROLEUM ENGINEERING. [J.C.C.]

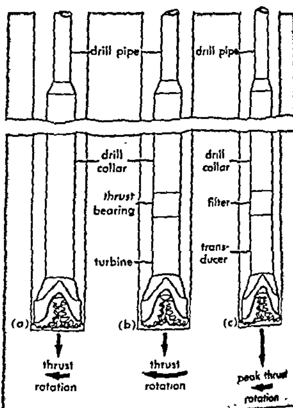
Bibliography: R L. Huntington, *Natural Gas and Natural Gasoline*, 1950; D. L. Katz et al., *Handbook of Natural Gas Engineering*, 1959; L. C. Uren, *Petroleum Production Engineering: Exploitation*, 3d ed., 1953.

Oil and gas well drilling

Certain features are common to all oil- and gas-well drilling. Rock and sand formations of the

be cased, or lined, ... seal out water, and to permit production of oil or gas.

Two principal drilling methods use a rotary tool or a cable tool The turbodrill method, which is a ... drilling, is widely used in the Soviet



Principles of (a) rotary drill, (b) turbodrill, and (c) vibratory drill. Arrows at bottom indicate relative importance of thrust and rotation.

90% of the 53,668 wells drilled. Because of its operating advantages, the older, cable tool method is still dominant in Ohio, Pennsylvania, Kentucky, West Virginia, and Michigan. Cable tools are used primarily in drilling shallow wells (usually above 5000 ft) through hard formations, and frequently to complete wells where low pressures are met in the subsurface formations.

Hole diameter is related to depth, and decreases as the hole goes deeper. For instance, a hole going down 20,000 ft. may be as wide as 24 in. for the first few hundred feet and may taper to less than 8⅝ in at the bottom

In addition to the methods mentioned, drilling techniques involving variations in the manner of conveying power to the bit, or cutting tool, are also under study. These experimental techniques include vibration drilling, sonic drilling, and rotary percussion drilling. See CABLE-TOOL DRILL; ROTARY TOOL DRILL, TURBODRILL. [A.L.P.]

Bibliography: L. C Uren, *Petroleum Production Engineering Oil Field Development*, 4th ed, 1956

Oil and gas wells

Holes that have been formed mechanically in the earth's crust, penetrating a commercial subterranean deposit or deposits of liquid or gaseous petroleum hydrocarbons or both, and so equipped that the formation product may be safely expelled or withdrawn.

The first commercial oil well was drilled at Titusville, Pennsylvania, in 1859, to a total depth of

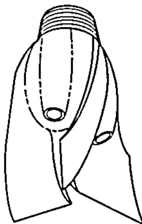


Fig 3 Sketch of a drag bit.



Fig 4 Photograph of a rock bit, a partial cut-away view (Hughes Tool Co)

run on logging cable It carries hollow bullets which are inserted in barrels and shot into the formation by means of explosive charges The bullets are positioned accurately at the proper depth and their firing is controlled at the surface The bullets and contained cores are pulled from the formation by retrieving cables or wires connected to the body of the instrument As many as 30 cores may be obtained on one trip into the hole

Core drilling. In this type of coring, the drilling bit is exchanged for a core barrel fitted with a core bit. The bit is designed to cut around the center of the hole and may be a roller-type or a diamond-type cutter head The core barrel receives the undrilled center portion and keeps it intact until it is pulled to the surface. As much as 50 ft of core may be taken at one time. *See* CORE DRILLING.

Oil-well fishing tools. During drilling, or production operations, or both, material or equipment may be accidentally left in the bore hole. This commonly results from fatigue failure of such equipment, usually a part or parts of the down hole drilling assembly (bit cutters, drill collars, or drill pipe). The material so left is called the "fish." The special tools or devices that are required for their recovery are consequently known as fishing tools. Most common of these tools is the releasing and circulating socket that is installed on the bottom of the intact drilling assembly. When run back into the bore hole, the socket will overshoot the top of the fish, and in the manner of a grapple, will rejoin the members, enabling the lost portion or fish to be recovered Conditions such as damage to the top of the fish or bad hole conditions may develop that would make further attempts at recovery uneconomical. In that event, the hole is either abandoned, or completed in a formation at a shallower depth, if one exists, or the fish is sidetracked, by directional drilling, before the main drilling is resumed

Directional drilling. In the normal course of drilling an oil or gas well, it is desirable to drill a hole as nearly as possible in a true vertical course. At other times, it is necessary to utilize a method of deviation from vertical called directional drilling The term directional drilling means controlling the course of a bore hole by using surface and subsurface instruments to reach a predetermined

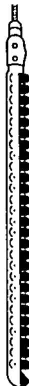


Fig 5 Sketch diagram of a side-wall sample gun.

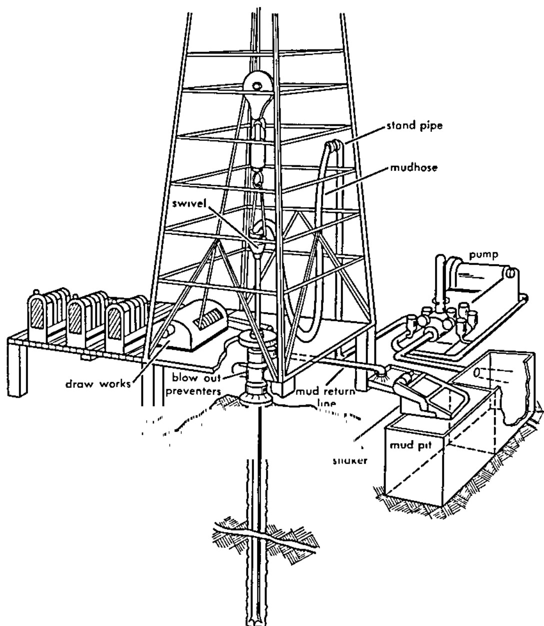


Fig. 2 Rotary rig and circulating system. In rotary drilling a bit is attached to the drilling end of a length of hollow steel pipe. As the pipe is spun by a rotary table, the bit bores into the earth, and more lengths of pipe are added as the hole goes deeper.

The blending, compounding, weighing, cleaning, and general treatment of the fluid require the attention of specially trained personnel.

Air or gas may be used as the circulating medium. Virtually the same equipment is used as with the circulating fluid, except that air compressors or a supply of high-pressure gas is used in lieu of the mud pumps Penetration rates are greater and drilling costs are lower with air or gas drilling, if used in selected areas of low-pressure formations in which little if any formation fluids are encountered prior to reaching the productive intervals, at which time circulating fluids are generally used (Fig 2).

Bits. The tool placed at the lower extremity of the drilling shaft is necessarily one that does the actual boring or drilling of the hole into the formations. This tool, called a bit, is designed to permit the removal of the entire cross section of the hole (cuttings) with the assistance of the circulating medium Bits are of numerous designs, but may be divided into two broad classes, drag bits and roller-rock bits.

Operational sampling and logging. At strategic intervals in the downward progress, formation samples are removed from the hole for surface inspection and analysis by the geologist for stratigraphic information or for evaluation of the zone for producing possibilities. One method of taking samples is by punching, as with a biscuit cutter; another, by drilling around, as with a hollow tube so as to preserve the core.

Punch cores. Biscuit-type punch cores are taken by mechanically forcing hollow tubes into the walls of the bore hole and retrieving the tube with a formation sample. Wire-line or drill-pipe and wire-line tools are used to accomplish this type of core taking. One such tool is a percussion type and is

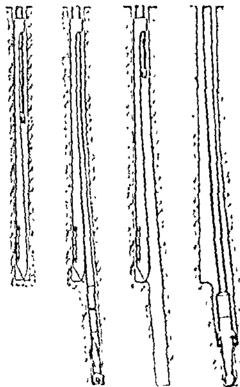


Fig 9 Operation of removable whipstock as a deflecting tool (Eastman Oil Well Survey Co)

setting All tools are run through these controls during all operational phases of the drilling and completion work. These valves are designed to close around the tools and effectively seal off against high pressures Valves positioned at suitable points in the blowout preventer assembly may then be manipulated to expel fluid or to receive fluid under pressure from rig or auxiliary pumps to bring the well under control

Casing, completion and connection. Various problems of well casing are related to modes of completing connection between the proper reservoirs and the surface

Casing. The casing for an oil or gas well is a steel tube, manufactured in various diameters, wall thicknesses, lengths, and steel alloys, selected to satisfy specific needs These lengths, called joints, are threaded and coupled so that they may be joined together in a continuous string in the well bore Properly placed and cemented in the hole, casing protects fresh-water reservoirs from contamination, supports unconsolidated rock formations, maintains natural separation of formations, aids in the prevention of blowouts and waste of reservoir energy, and acts as a conduit for receiving pipe of smaller diameters through which the well effluent may be produced to the surface under controlled conditions.

It may be necessary to set many strings of casing in one hole before reaching the objective. The determining factors are many, such as depth of hole, loss of circulation, high pressure formations, hole sloughing, and wearing out a string of casing while rotating drill pipe through it over a long period of time.

Cementing casing. Basically, ordinary portland cement is used in cementing casing. In order to obtain the protection and fulfil the purposes, it is imperative that each string of casing be securely sealed to the walls of the hole for at least some distance up from the bottom of the casing string. After casing is in place, the cement is pumped down the inside and up the outside to a predetermined height to occupy the space between the casing and the walls of the hole, thereby effecting the desired seal In the pumping and measuring process, plugs are used to separate the cement from other fluids to eliminate contamination; and the cement inside the casing is displaced with fluid. Oil- or gas-well cementing is not performed with the drilling equipment, but by an outside service company equipped with mobile, high pressure, mixing and pumping equipment and accessories operated by trained personnel

Completion techniques. Reservoir conditions, known to exist or later defined, determine the type of completion technique to be followed: (1) barefoot completion, (2) preperforated liner, or (3) casing set through and perforated.

1. Barefoot completion This is a type of completion frequently used when the character of the producing rock is such that it does not require supplemental support or screening, for example, in for-

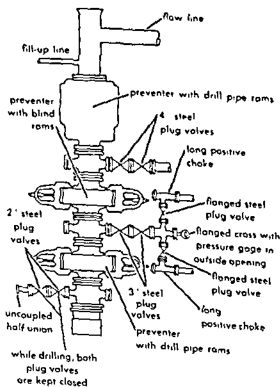


Fig 10 Blowout preventer, hookup diagram

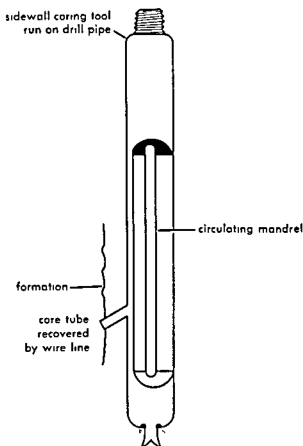


Fig. 6. A mechanical-type side-wall coring tool.

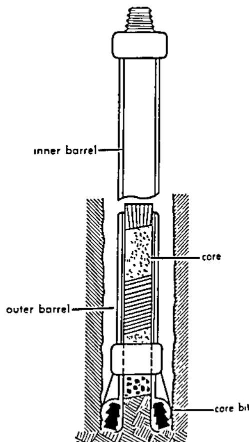


Fig 7. Drawing of core bit and core barrel

subsurface target. Holes may also be deflected to sidetrack obstructions (tools lost in the hole) or to reach a more strategic structural position In any case, mechanical means are employed to accomplish this by using directional recording instruments and deflecting tools to chart and control the downward course of the hole *See* PHOTOCLINOMETER

The most common of several types of deflecting tool is the whipstock, which is a wedge-shaped steel casting with a tapered or concave guide channel for the bit. In principle, the whipstock acts like a child's sliding board in deflecting the downward course The guide channel may be positioned at the desired compass point by way of a controlled method of orientation, using the earth's magnetic field as a point of reference.

High-pressure well control. Numerous factors are involved in high-pressure well control. High pressure in this sense means any pressure existing in a reservoir sufficient to expel a column of water contained within the well bore (exceeding the hydrobatic head for the depth). Initially, surface casing is set and cemented at such a depth that formations below the setting point will withstand pressures encountered during drilling operations should surface controls be closed.

Weighted fluids Adequately weighted fluids will develop a hydrostatic pressure sufficient to offset such formation pressures, but they may also be heavy enough to cause thieving of the fluid by other formations. In some areas, a change in weight of as little as $\frac{1}{10}$ lb/gal may result in loss of circula tion or blowout.

Blowout preventers. If the circulating fluid fails to contain the pressures, surface controls called blowout preventers are closed. These consist of valves that are previously installed above ground and fixed to the surface casing at the time of its

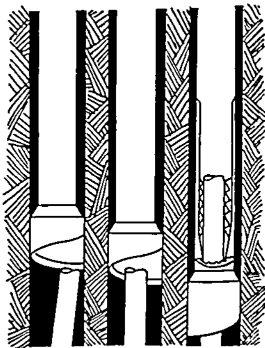


Fig. 8. Sketch of a socket (circu l g).

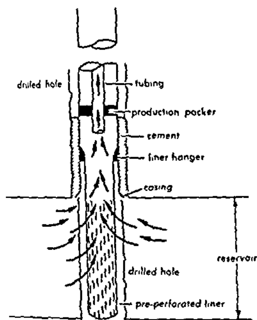


Fig 14 Liner-type completion, preperforated liner

well head assembly to which the so-called Christmas tree is attached

Christmas tree A manifold constructed of steel valves and fittings, placed on top of the casings protruding above the surface, is called a Christmas

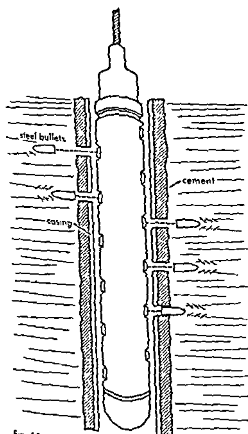


Fig 15 Gun perforator bullets from bullet-type gun

tree. Its purpose is to maintain the well under proper control, to receive the formation products under pressure, and to control the rate of daily production from the reservoir and direct it into a pipe line, generally at reduced pressures, to the oil gathering station.

Pumping unit. Relatively few oil wells are flowing from natural pressures. Most require secondary means of removing the reservoir product. The most common of several methods is the pumping unit. A walking beam is operated like a seesaw raising and lowering a plunger-type pump, set near the bottom of the hole. The rods between the walking beam and the pump are called sucker rods.

Multiple completion. An oil or gas well from which several separate horizons are individually, separately, and simultaneously produced is called a multiple completion. Such a completion is accomplished by the use of multiple-zone packers and separate tubing strings The producing zones are separated one from another by proper placement

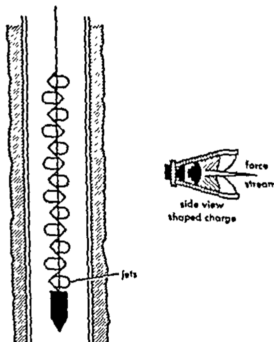


Fig 16 Diagram of jet-type perforator and jet-type guns

of packers in the well bore. An individual string of tubing is attached to each packer and extended to the surface where each is interconnected with the Christmas tree or flow assembly.

The several advantages of such a completion from a single well bore are increase in daily production, more efficient and economic utilization of a well bore with multiple reservoirs, increase in ultimate recovery, accurate measurement of product withdrawal from each reservoir, and elimination of mixing of products of different gravities and basic sediment and water content.

Water problems. The production of water in quantity from an oil or gas well ...

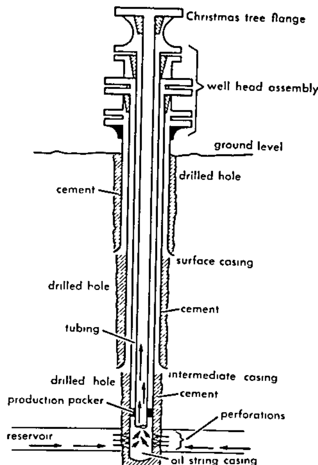


Fig. 11. Casing detail; casing strings in an oil well.

mations such as limestone, dolomite, or hard sandstone. With this method, the production casing is seated above the producing section in the conventional manner (see prior discussion of cementing) The casing is cleaned out by insertion of fluid separating plugs and drilled out through the casing and into the producing formation below. The formation contents enter the bore hole from the bare or unlined producing stratum or strata, hence the term barefoot.

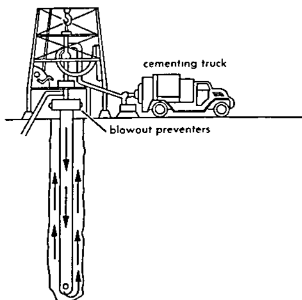


Fig 12. Sketch diagram of cementing; truck, equipment, and well job.

2. Liner-type completion. This type of completion is similar to a barefoot completion except that the open portion of the hole is cased with a preperforated section of casing called a liner. This liner is smaller in diameter than the casing previously set in the hole and is usually suspended from the upper casing near the bottom from a liner hanger. The hanger is attached to the top of the liner, and when it is set, it effects a seal between the liner and the casing. The purpose of the liner is to permit gas and liquids to enter the hole and screen out formation particles.

3. Gun perforating Gun perforating is a method of forming holes through the casing and into a formation from within a well bore. The two more popular methods are bullet perforating, as with a rifle, and jet perforating, as with a torch.

The gun is fitted around the outside with barrels containing the perforating medium Each barrel is wired to fire by remote control from the surface The gun is run into the hole on a wire line from a service company's shooting truck. The wire line serves to lower and raise the gun in and out of the hole, and when the gun is in position to be fired,

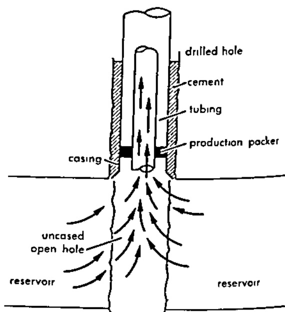


Fig 13 Diagram of barefoot completion

the operator sends an electric impulse down the line to trigger the gun The hole is thus formed through the steel casing, the cement sheath, and some inches into the reservoir rock, creating an entry through which the reservoir content enters the well bore. Guns of the bullet type are retrieved and reloaded (Fig. 15) Jet-type guns are expendable and disintegrate (Fig 16).

Tubing. A steel tube, the same as casing except that it is smaller in diameter, serves as a production flow line within the well. The tubing is run inside the casing and is either suspended in the hole or set on a production packer at or near the producing interval. The top of the tubing string terminates at the surface in a se ment in

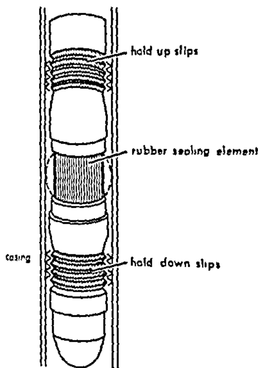


Fig 20 Sketch of a bridge plug

Water-exclusion plug back Simple water shut off jobs in shallow, deep, or high-pressure wells may also be performed in multizone wells in which the lower producing interval is depleted or the remaining recoverable reserves do not justify rehabilitation by placing a packer type plug (cork) above the interval, then producing formations that are already open or perforating additional intervals that may be present higher up the hole

Production-stimulation techniques. The initial testing or production history often indicates subnormal production rates signifying the necessity for remedial action. Any method designed to increase the production rate from a reservoir is defined as production stimulation. Three of the methods used are acidizing, fracturing, and employing explosives.

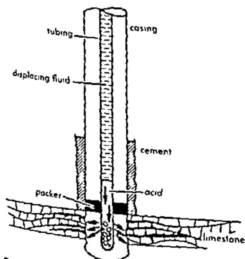


Fig 21 Outline of the acidizing process

Acidizing Varied volumes of hydrochloric acid are used in limestone and dolomite or other acid-soluble formations to dissolve the existing flow-channel walls and enlarge them. High-pressure equipment, pumps, and wellheads are necessary for satisfactory performance. Fast pumping speeds and acid inhibitors are used to alleviate corrosion of the well equipment.

Fracturing. Formation fracturing is a hydraulic process aimed at the parting of a desired section of formation Selected grades of sand or particles of other materials are added to the fracturing fluid in varied quantities. These particles pack and fill the fracture, acting as a propping agent to hold it

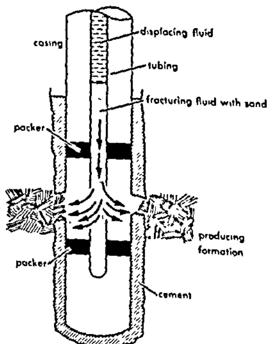


Fig 22 Outline of the fracturing process

open when the applied pressure is released. Such fractures increase the flow channels in size and number, improving the fluid flow characteristics of the reservoir rock. The particle carrying agent (fluid) is of considerable importance and is varied to fit particular demands. Some of the fluids used in this process are crude oil (sand oil fracturing), special refined oils (sand oil fracturing), water (river fracturing), acid (acid fracturing), and oil water, and chemical emulsion (emulsifracturing)

Explosives. The idea of stimulating production by use of explosives was first used in a well in Pennsylvania on January 21, 1865 The first torpedo consisted of 8 lb of gun powder contained in a metal tube, which was lowered into the well and

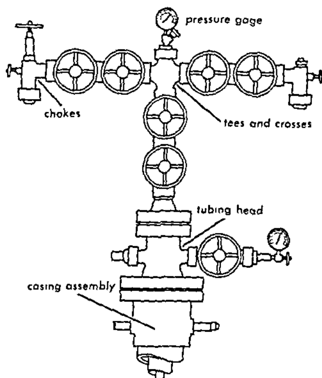


Fig 17 Typical layout of a Christmas tree.

nomical. Means are therefore provided for water exclusion.

Water-exclusion methods. Water exclusion may be effected by the application of cements or various types of plastic If it is determined that water is entering from the lower portion of a producing sand, in a relatively shallow, low-pressure well, a cement plug may be so placed in the bottom of the hole that it will cover the oil-water interface of the reservoir. This technique is called laying in a plug and may be accomplished by placing cement with a dump-bottom bailer on a wire line or by pumping cement down the drill pipe or tubing. For deeper, higher-pressure, or more troublesome wells, a squeeze method is used Squeeze cementing is the process of applying hydraulic pressure to force a cementing material into permeable space of an exposed formation or through openings in the casing or liner. In many conditions, cement, plastic, or diesel-oil cement may be squeezed into water, oil, or gas-bearing portions of a producing zone to eliminate excessive water, without sealing off the gas or oil. A few of the applications are repair of casing leaks; isolation of producing zones prior to perforating for production; remedial or secondary cementing to correct a defective condition, such as channeling or insufficient cement on a primary cement job; sealing off a low-pressure formation that engulfs oil and gas or drilling fluids; and abandonment of depleted producing zones to prevent migration of formation effluent and to reduce possibilities of contaminating other zones or wells.

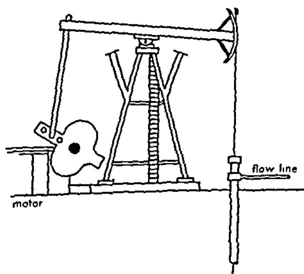


Fig. 18. Pumping unit diagram. These are most common if natural pressure is lacking for well flow.

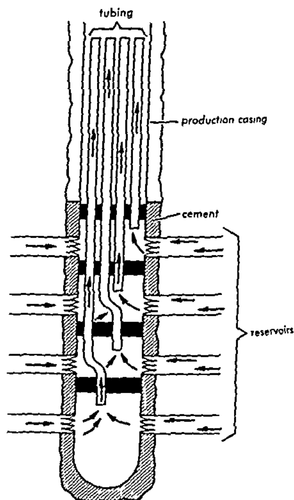
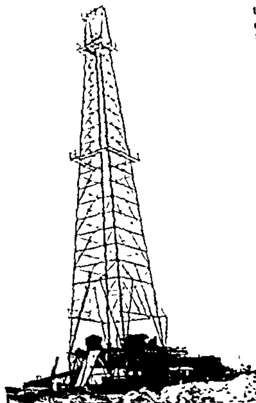
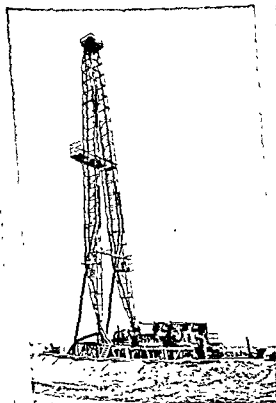


Fig 19. A representative multiple completion diagram

The squeeze-method tool is a packer-type device designed to isolate the point of entry between or below packing elements The tool is run into the hole on drill pipe or tubing and the cementing material is squeezed out between or below these confining elements into the problem area. The well is then recompleted. It may be necessary to drill the cement out of the hole and reperforate, depending upon the outcome of the job performed in the squeeze process.


Fig 25 Steel derrick, Lee C. Moore K type, 136 ft (Lee C Moore Corp)


Fig 26 Ideco full-view mast. (International Derrick and Equipment Company)

drilling methods are employed to drill and complete the well. The first combination cable-tool and rotary drilling rig was built in 1912. Essential differences in the three types of derrick are in the arrangement and design of the structural members used in the lowermost panel for bracing, and supports for the hoisting equipment installed in the sides of the derrick. One of the four derrick sides has a V opening with a minimum clear height of 23 ft 8 in. above the floor joists. This is the window opening which provides for pulling pipe and tools into the derrick.

Material. The Drake discovery well, completed August 27, 1859, America's first commercial oil well, was drilled with a boarded derrick unlike the later open framework wood derricks. This later type of wood derrick was used generally until 1892 when the first steel derrick made its appearance. Although some wood derricks are still in existence, used mostly for servicing wells, they had been fairly well superseded by the steel derrick for drilling purposes by the late 1920s and early 1930s. The preference for steel resulted from several factors, chief of which are (1) the fire resistance of steel, (2) the relatively smaller-size component members, which present much less area to wind pressure and therefore provide the steel derrick with a much greater wind load capacity, (3) the fact that steel derricks can be designed and fabricated with precision such that their safe working-load and wind-load capacities can be accurately determined, and (4) the fact that the steel derrick is easier to transport and can be erected and torn down in much less time. The wood derrick, however, may still have its application in localities in which suitable lumber and timbers are in abundance and available at low price.

Size. A derrick size is determined by its height, base square dimensions, and the clear opening of the water table or top of the derrick. The sizes and other general dimensions of derricks have been standardized by the American Petroleum Institute Division of Production Derrick heights are determined by the length of the hoisting tackle and tools in cable-tool drilling practice and by the number of individual lengths of pipe or joints in a stand of drill pipe and length of hoisting tackle in rotary drilling practice. Rotary drilling derricks for deep drilling when long stands of drill pipe are used are necessarily of greater height than cable-tool derricks.

Principal load members. The four legs and water-table beams are the principal load-carrying members of a derrick. For the wood derrick, the legs are built of 2×12 and 2×10 in timbers nailed together in the form of an L section. A structural-steel derrick has legs which are fabricated from structural-steel angles. Pipe has also been used for the legs of steel derricks. The leg of a derrick is built of several sections, and the ends of each leg section are squared and butted one against the other so that the load applied at the top or water table is transmitted directly through the legs.

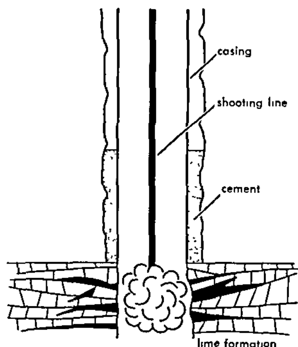


Fig. 23. Diagram of stimulating production through use of explosives.

detonated. Its more important function, in this shallow well, was to clear away paraffin, which was accomplished, but a decided increase in production was also accomplished. The method has since been improved with new explosives, firing mechanisms, and procedures, but the basic idea is the same, that is, to remove reservoir-blocking material from the reservoir face and to create fractures in the rock to increase production

Sand exclusion. Some reservoir rock is of an unconsolidated nature similar to beach sands, and after having been penetrated with a bore hole, it will slough, if unsupported, and has a tendency to flow with its formation fluids, resulting in plugging of the well bore and restriction or elimination of the entry of formation fluids. Several measures can be used to combat this condition, but no single measure can be used universally.

Screen liner. This type of liner is a segment of preslotted pipe wrapped with wire screen designed to screen out or retain outside the bore hole all except the fine particles that may be produced to the surface with the reservoir fluid. The original design is such that a calculated snug fit is obtained between the liner and the bore hole. The coarse screened-out particles form a secondary gravel pack between the liner and the hole, additionally supporting the formation and reducing sand incursion.

Sand consolidation. Sand consolidation is the result of successfully placing a binding material in the producing sand, and in effect gluing the sand grains together without completely destroying the porosity and permeability of the sand The binding material, generally a form of plastic, is forced into the sand through perforations in the casing. The purpose is to consolidate the sand around the well bore to eliminate sloughing and sand incursion

Prepack gravel liners. This type of liner (Fig. 24) is made by using a perforated section of steel pipe, over which has been fitted a tubular sleeve formed of an inner and outer screen of heavy wire mesh or perforated sheet steel and held in concentric relationship by spacers with gravel of proper size packed between the screens and sealed at both ends to retain the gravel. Set in the hole, it serves as a screen liner. *See* PETROLEUM RESERVOIR ENGINEERING. [H S BR]

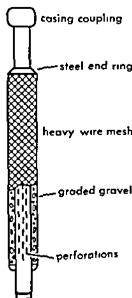


Fig. 24. Sketch of prepacked gravel liner.

OIL-WELL DERRICKS

The oil-well derrick is a framework or latticed tower of wood or steel, erected over the well for the purpose of hoisting and lowering pipe and tools in the well. The tower is composed of a lower framework or derrick substructure which supports the derrick floor and has working clearance beneath the floor for installation and operation of well-control equipment, and a superstructure generally supported by and bolted to the four heavy corners of the substructure, thereby forming a composite unit. In Fig. 25 a steel derrick and a wood derrick are shown.

A prefabricated steel mast, which is raised into operating position and lowered at the well as a unit with block and tackle, is superseding the conventional derrick in areas in which mobile rigs have application With such rigs, erection and moving costs are lower because of greater portability Portable masts of great height are fabricated in two or more sections by welding and the sections are then bolted or pinned together to make a complete unit assembly. Figure 26 illustrates an Ideco full-view mast and Fig 27 a Moore jackknife mast.

Derrick types. There are three types of conventional derricks: the cable-tool or standard derrick, the rotary derrick, and the combination cable tool and rotary derrick. The combination derrick permits a quick interchange from the rotary to the cable-tool system of drilling versa whe th

which failure of a steel derrick built according to American Petroleum Institute specifications may be expected is approximately twice its working load capacity.

Wind capacity. Like any structure exposed to wind pressure, a derrick must be designed so that it will resist the force of a wind of given velocity. Derricks intended for land operations only and where hurricane winds are unlikely or infrequent are designed for a wind pressure of 11.75 pounds per square foot (psf) acting on the exposed area of two directly opposite sides and a pipe setback standing within the derrick. When the derrick is to be used offshore where high-velocity winds can frequently occur, it is the practice to design for a wind pressure of 22.50 psf with a pipe setback and 52.90 psf without the pipe setback. These wind pressures correspond respectively to wind velocities of 54, 75, and 115 mph based on the formula $P = 0.004 V^2$, where P is unit wind load in pounds per square foot and V is actual wind velocity in miles per hour.

Foundation. Steel derricks are designed to be self sustaining when subjected to high wind pressure. However, the derrick must be securely anchored to foundation piers or corners of sufficient weight or having anchorage that will prevent the derrick's being overturned. Also, the pier footing must be large enough to distribute the maximum derrick load over the soil on which the structure rests without exceeding the safe bearing capacity of the soil.

Guying. When piers of insufficient weight are used, as is often the case, it is necessary for the derrick to be guyed adequately if it is to be prevented from overturning in a high wind. It is most important that the guys have the necessary strength and be attached to the derrick leg at which the bracing is also connected to avoid pulling the leg out of alignment. Also, the attachment must not damage the guy. Guy anchors should preferably be of the expanding or screw type, depending on the soil, and careful attention must be given their installation to fully develop the holding power of the anchor. The number and size of the guys will vary with the derrick size and the probable wind force that may act upon the derrick. In areas of high wind, it is general practice to guy a 136-ft steel derrick with three guys attached to each leg and spaced 21 ft apart.

[C.A.D.]

Bibliography. J. E. Brantly, *Rotary Drilling Handbook*, 4th ed., 1948; P. J. Jones, *Petroleum Production*, 1946; R. E. Sullivan, *Handbook of Oil and Gas Law*, 1955; L. C. Uren, *Petroleum Production Engineering: Oil Field Exploitation*, 3d ed., 1953, *Oil Field Development*, 4th ed., 1956; J. Zaba and W. T. Doherty, *Practical Petroleum Engineers' Handbook*, 4th ed., 1956.

Oil furnace

A combustion chamber in which oil is the heat-producing fuel. Fuel oils, having from 18,000 to 20,000 Btu/lb, which is equivalent to 140,000 to 155,000 Btu/gal, are supplied commercially. The lower flash point grades are used primarily in domestic and other furnaces without preheating. Grades having higher flash points are fired in burners equipped with preheaters.

The ease with which oil is transported, stored, handled, and fired gives it a special advantage in small installations. The fuel burns nearly completely so that, especially in a large furnace, combustible losses are negligible. *See* FUEL OIL; FURNACE (STEAM GENERATING).

Domestic oil furnaces with automatic thermostat control usually operate intermittently, being either off or operating at maximum capacity. The heat absorbing surfaces, especially the convection surface, should, therefore, be based more on maximum capacity than on average capacity if furnace efficiency is to be high. The combustion chamber should provide at least 1 ft³ for each 1.5–2 lb of fuel burned per hour. Gas velocity should be below 40 feet per second. The shape of the chamber should follow the outline of the flame. [F.H.R.]

Oil mining

The surface or subsurface excavation of petroleum-bearing sediments for subsequent removal of the oil by washing, flotation, or retorting treatments. Oil mining also includes recovery of oil by drainage from reservoir beds to mine shafts or other openings driven into the oil rock, or by drainage from the reservoir rock into mine openings driven outside the oil sand but connected with it by bore holes or mine wells.

Surface mining consists of strip or open-pit mining. It has been used primarily for the removal of oil shale or bituminous sands lying at or near the surface. Strip mining of shale is practiced in Sweden, Manchuria, and South Africa. Strip mining of bituminous sand is conducted in Canada.

Subsurface mining is used for the removal of oil sediments, oil shale, and Gilsonite. It is practiced in several European countries and in the United States. Some authorities consider this the best method to recover oil when oil sediments are involved, because virtually all of the oil is recovered.

European experience. Subsurface oil mining was used in the Pechelbronn oil field in Alsace, France, as early as 1735. This early mining involved the sinking of shafts to the reservoir rock, only 100–200 ft below the surface, and the excavation of the oil sand in short drifts driven from the shafts. These oil sands were hoisted to the surface and washed with boiling water to release the oil. The drifts were extended as far as natural ventilation would permit. When these limits were reached, the pillars were removed and the openings filled with waste.

This type of mining continued at Pechelbronn until 1866, when it was found that oil could be recovered from deeper, more prolific sands by letting it drain in place through mine openings, without removing the sands to the surface for treatment.

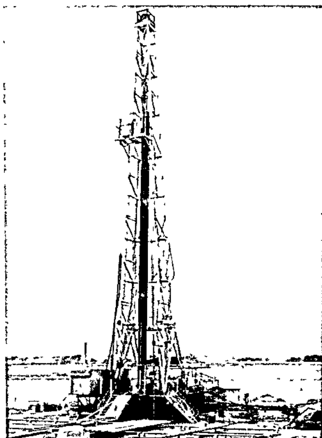


Fig 27. Moore jackknife mast (Lee C Moore Corp.)

Being a compression member, the load capacity of a derrick leg is dependent upon its slenderness ratio. This slenderness ratio is the unsupported length of the column or leg divided by the least radius of gyration of the area of cross section of the leg. A derrick may have its load capacity increased by reinforcement of the leg, provided the water-table beams are of a size to develop the reinforced capacity. Steel derricks are reinforced by clamping sections of pipe to the inside of the main legs and making the bottom and top sections bear firmly against stops to ensure that the reinforcements or relegs carry a proper share of the load Adjustment of the length of each releg is performed with a screw jack and the insertion of shims to obtain a firm bearing

Bracing. Horizontal members or girts are generally placed at uniform intervals to divide the leg into a series of short unsupported lengths The lowermost girt is attached to the legs at a distance of 10 ft above the derrick floor to provide working clearance for drilling and other well operations. Girts above the first are usually attached at intervals of 7 ft, forming 7-ft panels. Each panel is then braced with diagonal members, the composite bracing system consisting of derrick legs, girts, and diagonal braces, serving to counteract the transverse or horizontal forces to which the derrick is subjected.

The girts of a wood derrick are 2 × 12 and 2 × 10 in. timbers, and the inside braces are 2 × 10, 2 × 8, and 2 × 6 in. timbers Sway bracing is formed of 2 × 8 in timbers. In a structural-steel derrick, the girts and braces for the most part are structural angles, the size of the angle being selected according to the permissible maximum slenderness ratio of the member and the maximum load it must carry. The bracing of steel derricks has changed in style, and Kay- (or K-) type bracing which lends itself to efficient use of material has evolved.

Attachments. To provide safe access to all parts of the derrick and to facilitate working conditions at various levels, derricks are equipped with a ladder fastened to a side and extending from the bottom to the top of the derrick. Safety and working platforms installed at various levels are accessible from the ladder. Safety platforms are permanently located outside the derrick and are designated according to their particular height location. The platform at the top of the derrick is known as the crown platform or crow's-nest. A safety platform installed at a height of 80 ft is known as the quadruple platform, the height being equal to a stand of drill pipe made up of four 20-ft lengths of pipe. A double safety platform would be located at a height of 40 ft above the derrick floor.

Platforms installed inside the derrick at a working level are known as working platforms These include what are also called the monkey-board and pipe-stabbing board They are an accessory to the derrick and may be adjustable to working levels within prescribed limits. The top of the derrick is called the water table, and heavy-section wood timbers or bumpers, as they are called, on wood derricks, and steel structural wide-flange beams on steel derricks, are positioned on two directly opposite sides of the water-table opening. These beams support the drilling crown block and transmit to each of the four derrick legs the loads placed on the derrick in the course of drilling. Extending above the derrick top and across the water table opening is the gin pole which at its center supports a block arrangement that is used to raise the drilling crown block in the derrick and place it in position on the water-table beams.

Load capacity. Derricks vary in load capacity according to the size of the legs, and the water table beams are selected to develop this capacity The load capacity of a four-leg derrick is taken to be four times the capacity of a single leg acting as a column on the basis that each leg carries its proportionate share of the vertical load applied at the water table. This capacity, when determined in accordance with the American Petroleum Institute's allowable unit stresses for derricks, is displayed on a nameplate attached to the lowermost girt or section of derrick leg.

The legs of a derrick, however, are not equally loaded; therefore, the working load capacity will be less than the nameplate capacity. Because unequal forces are applied to the derrick legs, the nature and position of the loads acting on a derrick must be taken into consideration in order not to exceed the safe load capacity of a [illegible]k. The load at

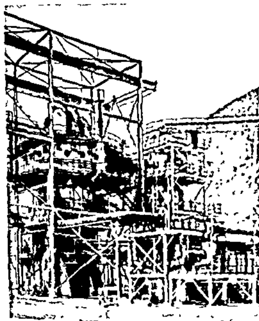


Fig 3 Close-up view of Union Oil Company of California shale-oil retort near Grand Valley, Colo Right part of the structure is portion of the systems for removing oil vapors that otherwise escape in the gas stream Men standing at ground level to the left indicate size of retort

retorted, and discarded for production of commercial quantities of oil. Various types of retort also have been tested in Colorado, but none is in commercial use.

Gilsonite. Gilsonite is a trade name, registered by the American Gilsonite Company, for a solid hydrocarbon found in the Uinta Basin of eastern Utah and western Colorado The American Gilsonite Company uses a subsurface wet-mining technique to extract about 700 tons of Gilsonite daily from its mine at Bonanza, Utah.

Conventional mining methods were found unsuitable for mass output of Gilsonite because it is friable and produces fine dust when so mined This dust can be highly explosive In the system now being used, tunnels are driven from the main shaft by means of water jetted through a ¼-in. nozzle under pressure of 2000 psi The stream of water penetrates tiny fissures and the ore falls to the bottom of the drift The drifts are cut on a rising grade of about 2.5° The ore is washed down to the main shaft where it is screened Particles of sizes smaller than ¾ in are pumped to the surface in a water stream; larger pieces are hoisted in buckets A long rotary drill with carbide-tipped teeth is used to remove ore that cannot be broken with water jets.

Gilsonite is moved through a pipeline in slurry form to a refinery at which the Gilsonite is dried and melted, and then heated to about 450°F The melted oil is fed to a coker and other processing units to make gasoline and other petroleum products

North American tar sands. Strip and open pit mining techniques have been established for the bituminous sands in the Athabasca region of Alberta, Canada, where the largest deposits of these sands are found. Bituminous sand, also known as asphalt rock, oil sand, and tar sand, is impregnated with heavily viscous oil that can be extracted by various methods, including hot- and cold water washing, solvents, centrifuges, and low-temperature distillation.

Once the overburden, which varies in thickness from a few feet to as much as 1800 ft, has been removed (where not too thick), the sand can be [illegible] er shovel [illegible] a small [illegible] st known underground mining methods could ever be used to recover more than 90% of the bituminous sand that lies under the heavier overburden, although it is known that some of the oil in these sands flows freely, leaving a possibility for some form of recovery by drainage

Cities Service Research and Development Company used a Krupp wheel excavator during test operations in the Mildred Lake area of the Athabasca region beginning in 1958. The excavator was designed to remove the oil sand from the bank in the quarry and carry it to a mobile extraction unit. The digging is accomplished by six 1.8-ft³ buckets mounted on a wheel The loosened oil sand is dropped on the conveyor belts, which transfer the oil sand to the vibrating grizzly on the extraction unit The excavator is capable of digging on a mining face 20 ft high by 30 ft wide by making three swinging cuts, each 6.6 ft high The maximum cutting depth is 11.5 in The mining capacity is 40 tons/hour The generator and diesel-engine drive, plus the necessary switchgear, starter batteries, fuel tank, and connecting lines, are mounted on a four-wheel trailer All operating parts are enclosed in a sheet-metal housing for weather protection, an important element in northerly loca-

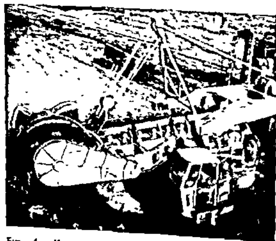


Fig 4. Krupp wheel excavator is used to remove bituminous sand near Mildred Lake, Athabasca region, Alberta, Canada (Cities Service Co.)

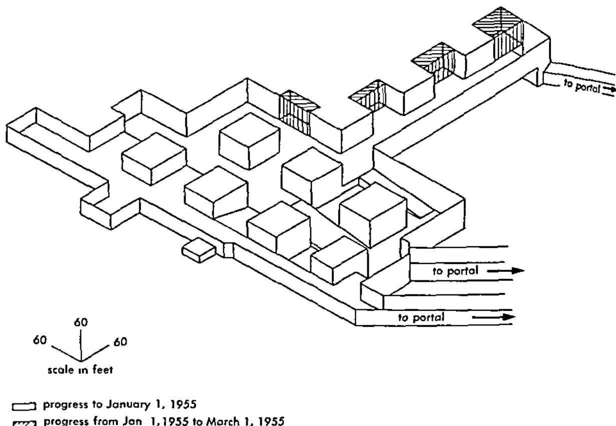


Fig. 1. Isometric drawing of U.S. Bureau of Mines experimental mine for oil shale at Rifle, Colo Developed by 39-ft top or advanced heading, a 34-ft following bench had just been initiated when a roof fall ended operations on Feb. 28, 1955. *(U.S. Bureau Mines Rept 5237, 1956)*

Subsurface mining of oil shale also goes back to the mid-nineteenth century in Scotland and France. It is not so widely practiced now because of its high cost as compared with those of usual oil production, particularly in the prolific fields of the Middle East.

United States oil shale mining. The U S. Bureau of Mines carried out an experimental mining and processing program at Rifle, Colo., between 1944 and 1955 in an effort to find economically feasible methods of producing oil shale.

One of the more important phases of this experimental program was a large-scale mine dug into what is known as the Mahagony Ledge, a rich oil shale stratum that is flat and strong, making it favorable for mining. This stratum lies under an average of about 1000 ft of overburden and is 70–90 ft thick.

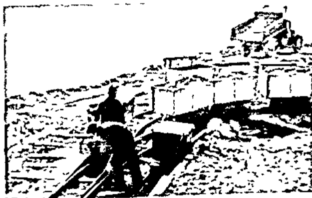


Fig 2. Mine locomotive and cars removing shale from U.S Bureau of Mines shale mine west of Rifle, Colo Segments of a light streak at left in middle ground are the Colorado River, nearly 3000 ft below. *(From R Fleming, U.S. Bureau of Mines)*

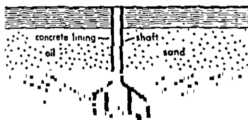
The Bureau of Mines adopted the room-and pillar system of mining, advancing into the 70-ft ledge face in two benches The mine roof was supported by 60-ft pillars staggered at 60-ft intervals and supplemented by iron roof bolts 6 ft long.

Multiple rotary drills mounted on trucks made holes in which dynamite was placed to shatter the shale, the shale was then removed from the mine by electric locomotive and cars.

The experimental mining program ended in February, 1955, when a roof fall occurred. Despite this occurrence, however, the Bureau is convinced that the room-and-pillar method used in coal, salt and limestone mines is feasible for shale oil mining in Colorado

Oil shale does not contain oil, as such. Draining methods, therefore, are not applicable. It does however, contain an organic substance known as kerogen. This substance decomposes and gives off a heavy, oily vapor when it is heated above 700°F in retorts. When condensed, this vapor becomes a viscous black liquid called shale oil, which resembles ordinary crude but has several significant differences.

Colorado's Mahagony Ledge yields an average of about 30 gal of oil per ton This means that large amounts of oil shale must [illegible] d, transported



g 7 Sketch of the Wright system of draining oil ınds by a series of bore holes drilled from a mine ıoft (After L C Uren, *Petroleum Production Engineer-ıg, Oil Field Exploitation*, 3d ed, McGraw-Hill, 1953)

ırough the productive strata, followed by long, lanting holes drilled radially in all directions from ıe shaft bottom into the oil sands These bore oles can be as long as ½ mile, and 2–5 in. in ıameter. This method was used in shallow oil ands in southeastern Ohio and in Pennsylvania.

Selection of method. Use of the various mining ıethods depends largely on conditions found They ıre not usable where high gas pressures may be ıresent Water can likewise be a major problem. n some cases oil sands are poorly consolidated, o that they flow under pressure, making mine sup-ıort difficult, if not impossible; in addition, these ıands tend to mix with the oil

Where subsurface conditions are good, that is, ıhere gas pressure is not high, water is not ex-ıessive, and sand conditions are favorable, mining ıan be used, provided costs can be kept at a com-ıetitive level *See* OIL AND GAS WELLS [A L P]

Bibliography F. L. Hartley and C. S Brinegar, Oil shale and bituminous sand, *Energy Resources Conference*, 1956; *Synthetic Liquid Fuels, Annual Report of the Secretary of the Interior for 1955, Part II, Oil From Shale*, U S Bureau Mines Rept. 5237, 1956; L. C. Uren, *Petroleum Production Engineering, Oil Field Exploitation*, 3d ed, 1953

Oil sand

A bituminous sand, asphalt rock, or tar sand, a loose sand or sandstone impregnated with very viscous oil *See* ASPHALT AND ASPHALTITE Among the best known oil sand deposits are those near Vernal, Utah, and those adjacent to the Athabaska River surrounding Fort McMurray in Alberta, Canada

The Athabaska tar sand is the largest deposit of this type known, having been estimated to contain from 100,000,000,000 to 300,000,000,000 or possibly even 500,000...

... made that it is a naturally stripped petroleum reservoir (fossil oil field), but other work has indicated that it might be of nonmarine origin and not directly related to petroleum

A number of attempts have been made to recover the oil from oil sands by solvent extraction, hot or cold water washing, retorting, or centrifugal separation techniques Once isolated, the oil can be refined by conventional processes.

The total reserve of oil in major oil sand formations of the United States has been estimated to be 2,000,000,000–3,000,000,000 barrels. [I A B]

Oil shale

A fine-grained, usually dark-colored sedimentary rock containing complex organic matter which, on heating, decomposes to yield oil (*see* KEROGEN). In its purest sense oil shale is a commercial term used to designate those shales that yield commercial quantities of oil (over 10–15 gal/ton).

Origin and composition. Oil shales may be either marine or lacustrine in origin. Each shale represents the slow accumulation of inorganic sediment together with the organic debris contributed by aquatic floral and faunal assemblages. A major contribution to the organic constituents consists of the pollen and plant fragments carried into the sedimentary basin by wind or streams. Following deposition, the organic ...

... spores have contributed the major portion of the organic detritus, the shale may take on the properties of an impure cannel coal. Humic substances derived from land plants may contribute substantially to the organic constituents of certain shales.

Most oil shales contain only small percentages (about 5%) of organic matter extractable under the usual laboratory conditions. Most organic matter is present in the form of high molecular weight, insoluble complex material.

Quartz and silicate minerals, especially clays ...

... minerals during the diagenesis of the sediment to produce the shale (*see* DIAGENESIS). Pyrite or marcasite, along with minor amounts of feldspars, is usually present Feldspars may also be formed in situ in a shale. The massive Green River oil shale of Colorado, Utah, and Wyoming is in reality largely composed of dolomite and calcite in addition to clays and other minerals commonly associated with shales

Many oil shales are thinly laminated (Fig 1), and parts of the Green River shale are known as paper shale; other shales exhibit a subconchoidal or conchoidal fracture

Oil-forming shales. Oil shales occur in many parts of the world, usually as relatively thin (several hundred feet) widespread formations. Some of the marine carbonaceous shales, such as the Miocene Monterey shale of California, are thought to have been source beds in which petroleum was formed. In others, such as the Chattanooga shale of the southeastern United States or the Swedish

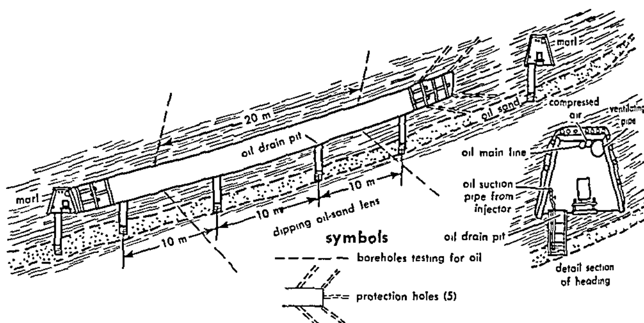


Fig. 5. Sketches illustrate method of conducting oil mining operations in the Pechelbronn field, Alsace, France (*After G S Rice, U.S Bureau of Mines*)

tions. The trailer also is designed for easy movement across the mining face.

Mine draining of reservoirs. Drainage of oil in place, although less efficient than actual mining, is cheaper and can be carried out with much less risk. This method has been practiced for years in the Pechelbronn field and in the Wietze field, near Hanover, Germany. A similar venture was successfully used in Ventura County, Calif, in 1866, as well as in other states since then.

In the Pechelbronn field, shafts were sunk to the level of the oil reservoir and drifts driven into the oil rock. Large drainage surfaces were available from wells drilled from the surface in previous years and later abandoned. Unrecovered oil drained down these abandoned wells into the mine workings, where it was gathered in underground storage sumps and then pumped to the surface All

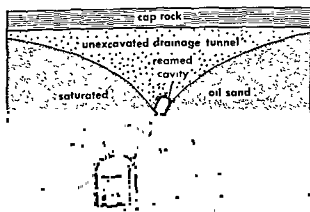


Fig. 6. Diagrammatic illustration of unexcavated drainage tunnels in connection with the Ranney system of oil mining (*After L. C. Uren, Petroleum Production Engineering, Oil Field Exploitation, 3d ed., McGraw-Hill, 1953*)

the tunnels driven from the shafts directly through the oil rocks afforded drainage surface too This oil accumulated in a trench cut in the floor of each tunnel. Oil flowed by gravity down these trenches to the central storage sump. This method of recovery was so efficient that Pechelbronn became one of the largest mining projects in the world, with more than 200 miles of tunnels in three separate mines. In addition, hundreds of miles of 1.5–2-in bore holes have been drilled for exploratory purposes and to check for high-pressure gas or oil pockets.

In another mining method that became more prevalent after 1925 at Pechelbronn, tunnels were driven through the impermeable rock cap above the oil sand, and at 33-ft intervals, 20–30-in square pits were dug through the tunnel floor to the oil sands, usually 6–8 ft below The oil drains into these pits, which are timber lined, and is lifted periodically by a pneumatic system into a pipeline extending to surface tanks.

One variation of this method, known as the Ranney oil-mining system, involves driving mine galleries in impermeable strata above or below the oil sands These galleries are driven from shafts communicating with the surface. Holes are drilled at short intervals along these galleries and gas or oil is withdrawn through pipes sealed into the bore holes. These mine wells communicate with a system of pipe drains in the tunnels, leading to separate tanks from which the oil is pumped to the surface, or the gas permitted to flow to the surface. There are several other methods which are similar to the Ranney system in that they also provide for oil drainage through holes drilled from mine tunnels outside the oil beds.

Another method proposed for mining oil from partially drained sands i lling a shaft

To obviate mining and disposal problems, the Swedish Shale Oil Co in 1944 applied the Ljundstrom in situ method for the distillation of oil. Electrical heating elements are inserted into a hexagonal pattern of drill holes sunk vertically through the shale bed, and the shale is gradually heated to 400°C over a period of five months The approximate yield per square meter of the field is gasoline, 515 liters; kerosine, 160 liters; heating oil, 350 liters; liquefiable gas, 80 liters; sulfur-free gas, 650 liters [illegible]

[illegible]

duction method. The oil, which contains paraffinic, olefinic, naphthenic, and aromatic constituents, may have a pour point ranging from 90°F (semisolid) to below 5°F. There is usually a preponderance of straight-chain hydrocarbons (approximately 50%) among the identifiable compounds in the distillate; sulfur compounds are mainly thiophenic Conversion of most of the distillate to motor fuel can be carried out by thermal or catalytic cracking, hydrogenation, or by other processes common in the petroleum industry See PETROLEUM PROCESSING

Shale oil is a source not only of fuel but also of a paraffin type wax, tar acids, and tar bases Tar acids consist of phenol, cresols, xylenols, and carboxylic acids; tar bases contain homologs of pyridine and quinoline.

The reserves of shale oil have been conservatively estimated to be 960,000,000,000 bbl for the Green River formation alone, the rich part of the formation, the Mahogany Ledge, has been estimated to contain 90,000,000,000 bbl. The enormity of this reserve is fully grasped when it is realized that the entire world has produced only 90,000,000,000 bbl of crude oil between 1859 and 1957 See MINERAL FUEL AREAS. [L A B]

Bibliography: F. L. Hartley and C. S Brinegar, Oil shale and bituminous sand, *Sci. Monthly*, 84-275-289, 1957; G Sell (ed), *Oil Shale and Cannel Coal*, vol 2, 1951.

Oil-field model

A small scale and commonly simplified replica of subsurface conditions of interest and value in petroleum prospecting and oil field development. The term model has been applied by geologists to a simplified diagram and by mathematical physicists to a formal analysis with special boundary conditions and related attributes These and somewhat more complex physical models have value in transmitting concepts and relationships to the nonspecialist, but they are generally designed by geologists and petroleum engineers to aid investigation of a particular problem To do so, the model is either scaled to size, shape, and like attributes, or according to forces upon it, relative to its archetype

Size-scaled models. These are commonly models of ore and mineral bodies, and of those blocks on which surface topography and surface geology are shown on top and geologic cross sections on the sides, also to scale. Peg models are constructed for oil fields with wells shown as pegs or wires. Cut out block diagrams are also used Models of this kind are found in many geological museums and in mining companies' offices.

Forces and movement models. This type contains mobile material to simulate by its movement a movement that has taken place or will take place in the prototype Transparent solid plastics have been incorporated in mobile parts of models for enhanced internal visibility and to obtain photoelastic data. Such models are also used to make graphic representation of changes recorded from observations with geophysical instruments, for example, from microinstruments used to survey physical response of a small mass in an artificial field Force and movement models have come into use for study of movements of earth materials By this, past changes resulting in present features can be made as dimensionally credible experiments in periods tremendously short as compared with geologic time. Models with mobility are of three principal types.

Fluid or fluids moving in porous medium. The models for study of fluid movement through a geometrically stable medium have important applications to hydrologic and petroleum engineering. The hydrologic case is mostly of an interface underground between fresh water and salt water caused by extraction or by injection of fresh water through wells Where the interface owes its position to fresh-water–salt-water patterns found in many coastal zones and along many shorelines, the setting is difficult to describe without a model. Such demonstrations have added to understanding of water conditions in California's Santa Clara Valley and other critical water areas.

The petroleum-engineering application is to movement of fluid in respect to another kind of interface underground. This is between petroleum and natural gas, or both, and salt water, and especially when petroleum is being extracted through wells and salt water simultaneously injected through other wells into the same porous stratum, so that the important flow is radial, that is, two-dimensional

The model used involves scaling down from the actual distance between wells in the oil field under study, and also the device of an analog for the fluids whereby the pressure distribution in steady-state porous flow is exactly the same as the potential distribution in an electrical conducting medium

Uniform high-viscosity materials. The second geologic realm where models are used extensively is that of the movement of large segments of the earth made of materials of essentially uniform high viscosity, comprising nearly all rock species in the earth's crust. This has little direct bearing on oil-field models

Adjacent materials of diverse viscosities Among the movements of the earth which are of interest to geologists are those involving [illegible]

alum shale, there is little or no associated crude oil and no indication that these shales were source beds. Little is known about the differences between shales that are or are not likely source beds, but these differences may be related to a great extent to types of organic matter present at the time of deposition.

The reducing conditions necessary for the preservation of organic matter in a shale-forming environment are conducive to the precipitation of available trace elements or the formation of certain minerals. Pyrite (FeS_2) occurs in large amounts in most shales The Kupferschiefer of Mansfeld, Germany, contains an unusually high content of copper, and the Swedish alum shale has been exploited for its uranium content (about 0.04%). This shale also contains in its upper units small lenticular masses of organic matter, called Kolm, that contain up to 0 5% uranium. The Chattanooga shale, containing about 0.006% uranium, has been extensively studied as a potential low-grade source of this element.

Current economic factors call for a shale to yield approximately 15 gal/ton of oil before it can be profitably exploited. During World War II, the Swedish shale, yielding about 12 gal/ton, was retorted as a source of oil. The Eocene Green River shale averages 15 gal/ton, but selected segments of the formation, such as the Mahogany Ledge, average about 30 gal/ton Only this latter part of the Green River formation is considered to be of present economic significance. Most other shales of the United States yield an average of only 10–12 gal/ton of oil, making them commercially unattractive. Other shales, with indicated oil yields, are shown in the accompanying list:

Source	*Oil yield, gal/ton (approx)*
Scotland	22
France	10–17
Estonia (kukursite)	48–86
Australia (torbanite)	80–180
Manchuria	15
Sweden	12–15
Russia	35–50
Germany	12
England	40
South Africa (torbanite)	106
Canada	22

As indicated in the list, those shales with high yields of oil have been given special names indicating them to be somewhat out of the ordinary in appearance, origin, and composition. *See* TORBANITE.

Recovery of oil. Oil is generally recovered from shale by techniques in which the ground rock is fed into a retort, the distillate is condensed, and the gases are vented Many retorts operate on a batch basis and employ combustion of part of the organic matter of the shale to provide the heat necessary for distillation The usual retort is able to handle 40–50 tons of shale daily. Continuous-type retorts have also been used in which raw shale is fed into one end of a heated tunnel and expended shale is withdrawn from the other end.



Fig. 1. Thinly laminated oil shale. (*Union Oil Co of California*)

Recovery of oil from shale is dependent to a great extent upon the development of satisfactory mining methods for the shale and upon means for the disposal of large quantities of spent shale (hot shale ash). The U.S. Bureau of Mines has operated a demonstration plant at Rifle, Colo, for a number of years and has carried out considerable research on mining and retorting techniques for use with the Green River shale The Union Oil Co, operating a pilot plant (Fig 2) in Grand Valley, Colo, developed a daily throughput of 1200 tons of Green River shale leading to production of 750 bbl/day of oil.

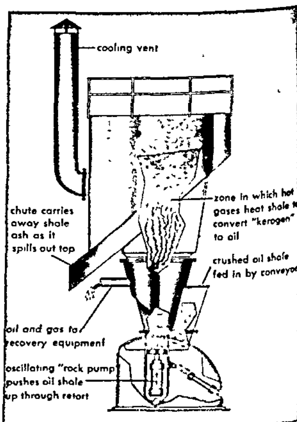


Fig 2. Pilot-plant model of oil shale retort Shale is fed at bottom and retorted in center section (*Union Oil Co of California*)

mently the

the water on minerals in the reservoir rock and , the mechanical equipment employed on the oject Waters which exert a corrosive action on e lines and pumps or which tend to plug up the ty sand are not suitable for water-flooding operans

Oil field water composition may be an important ictor in the determination of the source of water ı oil wells which have leaky casings or in identiıng and correlating reservoirs in multi-pay oil ools, particularly in those containing lenticular and bodies.

Industrial wastes, including mineralized water roduced with oil, may be disposed of in underground reservoirs. Between the zone of potable water and the horizon of commercial brines, there commonly are rock formations, the waters of which contain chemicals in amounts sufficient to make the waters unsuitable for domestic, municipal, industrial, and livestock consumption, but not in sufficient quantity to be considered as a source for recovery of chemicals Provided there is sufficient porosity and permeability, these rock formations could receive industrial wastes which would contaminate surface streams and shallow, fresh groundwater horizons into which they might otherwise be discharged *See* GEOPHYSICAL EXPLORATION; PETROLEUM GEOLOGY, WELL LOGGING (MINERAL)

[P M.]

Okra

A warm season annual (*Hibiscus esculentus*) of Ethiopian origin Okra, also called gumbo, is grown for its immature pods which are generally used for preparing soups, but are also eaten as a freshly cooked vegetable. It is a member of the plant order Malvales and is related to cotton. Propagation is by seed Popular varieties are Clemson Spineless and Green Velvet. Okra is sensitive to low temperatures; commercial production in the United States is primarily in the South. Harvesting begins when the pods are 3–4 in. long, usually 50–60 days after planting. Georgia, Florida, and and Louisiana are important producing states. *See* COTTON, MALVALES; VEGETABLE GROWING [H.J.C.]



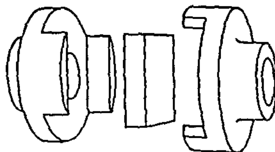
Okra, *Hibiscus esculentus* (a) Pods (fruits) (b) Leaf (From L H Bailey, ed, *The Standard Cyclopedia of Horticulture*, vol 2, Macmillan, 1937)

Oldham's coupling

A flexible coupling that permits two slightly misaligned shafts to be joined In the conventional form of Oldham's coupling, as illustrated, each shaft end is fitted with a hub having a smooth slot or groove A floating member with orthogonal ridges or tongues mates with the two hubs. The two degrees of freedom thus provided enable the coupling to accommodate to axis displacement, but not to angular displacement A similar action is provided by a square block mating with widely slotted hubs Adequate lubrication is essential to efficient operation *See* COUPLING; UNIVERSAL JOINT. [J.J.R]



conventional Oldham's coupling

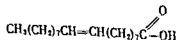


modern coupling with Oldham principle

Parallel displaced rotating shafts are joined by a coupling with a floating center member.

Oleate

A salt or ester of oleic acid



$$CH_3(CH_2)_7CH{=}CH(CH_2)_7C({=}O){-}OH$$

in which the acid hydrogen is replaced by a metal or an organic radical. Oleates occur in nature chiefly as the glyceryl ester, found in substantial amounts in animal and vegetable fats. A few of the simpler esters have commercial applications in

soft layers between relatively hard ones. This is the realm where diapir structures are inferred to develop (*see* DIAPIRIC STRUCTURES) and also their subform, the salt dome (*see* SALT DOME). Because of the economic importance of salt domes and their associated oil traps, investigations with models have become widespread.

L. L. Nettleton produced a model for salt-dome formation using two viscous fluids of different density to emphasize the role of the lower density of salt to that of adjacent sediments in the process of salt-dome movements. M. B. Dobrin made a parallel mathematical analysis T. J. Parker and A. N. MacDowell extended salt-dome studies by using materials with higher yield points than Nettleton's first model, and their model showed fractures above and around the "salt" analog in their model.

Also in both the prototype and the model, boundary conditions modify vectors nearby. This seems particularly hard to handle in cases of hydrologic model study such as at the University of California.

Models versus mathematical analysis. There is a possible alternative to the use of models—mathematical analysis, which is more attractive now that computing machines are available. Dobrin used this method for the case of Nettleton's model investigation of salt dome However, mathematical analysis is as yet a subject of controversy among physicists in many simple geologic settings; models are presently available and their value is acknowledged.

Models are probably more helpful in educating the nonspecialist in geology, they present the earth features in a readily visible fashion easier to grasp than mathematical analysis. [P.WE.]

Bibliography: M B. Dobrin, Some quantitative experiments on a fluid salt-dome model and their geological implications, *Trans. Am. Geophys. Union*, 22:528–542, 1941, H. E. McKinstry et al, *Mining Geology*, 1948; L. L Nettleton, Recent experimental and geophysical evidence of mechanics of salt-dome formation, *Bull. Am Assoc. Petrol. Geologists*, 27(1):51–63, 1943; A. E. Scheidegger, *Principles of Geodynamics*, vol. 1, 1958; D K. Todd, *Ground Water Hydrology*, 1959.

Oil-field waters

Waters of varying mineral content which are found associated with petroleum and natural gas or have been encountered in the search for oil and gas. They are also called oil-field brines, or brines They include a variety of underground waters, usually deeply buried, and have a relatively high content of dissolved mineral matter. These waters may be (1) present in the pore space of the reservoir rock with the oil or gas, (2) separated by gravity from the oil and gas and thus lying below it, (3) at the edge of the oil or gas accumulation, or (4) in rock formations which are barren of oil and gas. Brines are commonly defined as water containing high concentrations of dissolved salts. Potable or fresh waters usually *are not* considered oil-field waters but may be encountered, generally at shallow depths, in areas where oil and gas are produced.

Oil-field waters or oil-field brines differ widely in composition and concentration They may differ from one geologic province to another, from one formation to another within a given geologic province, or from one part of a specific geologic horizon to another. They range from slightly salty water with 1000–3000 parts of dissolved substances in 1,000,000 parts of solution to very nearly saturated brines with dissolved mineral content of more than 270,000 parts per million (ppm).

The most common and abundant mineral found in oil-field waters is sodium chloride, or common table salt. Calcium chloride is next in order of abundance Carbonates, bicarbonates, sulfates, and the chlorides of magnesium and potassium are present in lesser quantities. In addition to the above mentioned salts, salts of bromine and iodine also are found. Traces of strontium, boron, copper, manganese, silver, tin, vanadium, and iron have been reported. Barium has been reported in many of the Paleozoic brines of the Appalachian region The commercial value of a brine depends upon the concentration of salts, purity of the products to be recovered, and value and practicability of by product recovery. Concentrations less than 200,000 ppm are seldom of commercial interest.

Classified genetically, oil-field waters are generally considered connate; that is, they are sea waters which (presumably) originally filled the pore spaces of the rock in which they are now confined However, few analyses of these waters correspond to present day sea water, thus indicating *some mixing and modification* since confinement. Dilute solutions suggest that rainwater has percolated into the rocks along bedding planes, fractures, faults, and other permeable zones Presence of carbonates, bicarbonates, and sulfates in an oil field water further suggests that at least some of the water had its origin at the surface Concentrations of dissolved solids greater than that of modern sea water suggest partial evaporation of the water or addition of soluble salts from the adjacent or enclosing rocks.

Waters in most sedimentary rocks increase in mineral concentration with depth. This increase may be due to the fact that since salt water is heavier than fresh water, the more dense solution will eventually find a position as low as possible in the aquifer An additional factor would be the longer exposure of the deeper waters to the mineral-bearing rocks. Exceptions have been noted and probably are due to the presence of larger quantities of soluble salts in some geological formations than in others.

Probably the most important geological use of oil field water analyses is their application to the quantitative interpretation of electrical and neutron well logs, particularly the recently developed micrologs. In order to compute the connate water saturation of a formation in a quantitative manner from electrical data it is necessary to know with accuracy the connate wate [illegible]ty.

Fauna. Oligocene mammals are intermediate between the archaic forms of the Paleocene and Eocene and the modern forms of the Miocene and Pliocene. All of the families of mammals now living have Oligocene representatives, and very few of the primitive groups, such as the rhinoceroslike titanotheres, survived The earliest elephants and manlike apes are found in Oligocene deposits in Europe and Africa Tropical marine faunas became more closely restricted to the equatorial region than they were in Eocene time The Mediterranean-eastern Atlantic fauna became progressively more differentiated from the western Atlantic (American) fauna during the Oligocene See PALEOBOTANY; PALEONTOLOGY

[A H CH , G E M]

Oligochaeta

A class of the phylum Annelida including worms such as the earthworms These animals exhibit both external and internal segmentation Except in the family Branchiobdellidae, they possess setae which are not borne on parapodia Oligochaetes are hermaphroditic The gonads are few in number and situated in the anterior part of the body, the male gonads being anterior to the female gonads. The gametes are discharged through special ducts, the oviducts and sperm ducts. A clitellum is present at maturity, in which the eggs are deposited in a cocoon There is no larval stage during development

The oligochaetes are primarily fresh-water and burrowing terrestrial animals A few are marine and several species occur in the intertidal zone Earthworms are the familiar members of the class They range in size from aquatic species which are 0.5 mm long to giant earthworms several feet long There are 21 families with over 3000 species

Body form and anatomy. Oligochaetes are cylindrical, elongated animals with the anterior mouth usually overhung by a fleshy lobe, the prostomium and the anus terminal, except in a few epizoic species The body plan is that of a tube within a tube Externally, the segments are marked by furrows The setae or bristles, usually a pair of ventrolateral bundles, are borne on most segments Other external features are the pores of the reproductive systems opening on certain segments, the openings of the nephridia and in many earthworms dorsal pores which open externally from the coelom. Some aquatic species have extensions of the posterior part of the body which function as gills

Body wall The body wall consists of a simple columnar epithelium which secretes a thin cuticle, muscle layers, and a lining peritoneum The epithelium contains interspersed glandular and sensory cells The muscles are arranged in an outer circular and an inner longitudinal layer Movement in terrestrial forms is accomplished by alternate contractions of these layers which produce a peristaltic motion in which the setae aid forward progression by gripping the substrate Other smaller muscles

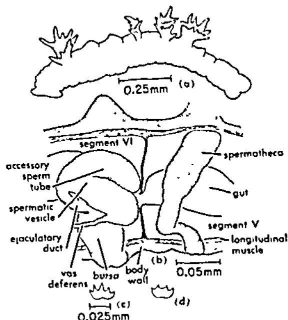


Fig 1 Branchiobdellidae. (a) *Pterodrilus alcicornus*, lateral view (b) Segments V and VI of *P. alcicornus*, showing reproductive structures (c) Upper jaw (d) Lower jaw of *P alcicornus*

protract and retract the setae and are associated with the intestine and reproductive organs.

Coelom The coelom is filled with a fluid and segmentally partitioned in most cases by the septa The alimentary canal passes through this series of spaces and is composed of a buccal cavity, pharynx, esophagus, and the intestine proper In the earthworms, a crop and a gizzard are often present. Various pharyngeal and esophageal glands may be associated with the alimentary canal In many of the larger oligochaetes the lining epithelium, composed of glandular and absorptive cells, projects downward from the dorsum as the typhlosole which is a longitudinal fold into the lumen of the intestine. There are thin muscle layers in the wall of the intestine and it is invested with a layer of modified peritoneum, the chloragogen, which has an assimilative and perhaps excretory function. The food of most oligochaetes is plant material ingested by a sucking action of the pharynx. Only the members of the family Branchiobdellidae are provided with pharyngeal jaws and they and some other aquatic forms are carnivorous; a few others are parasitic

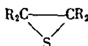
Nephridia and excretion Excretion in the oligochaetes is accomplished by segmentally arranged paired nephridia. Typically each is composed of a funnel, the nephrostome which opens into the coelom and a variously coiled tube which discharges ventrolaterally to the outside Differences in the number of cells which compose the lip of the nephrostome are considered to be of taxonomic importance

Respiration The exchange of respiratory gases is in general a function of the body surface Some aquatic forms, *Dero*, have anal gills and some other

textiles, leather, and cosmetics. Alkali metal oleates are water-soluble and, with the similar stearates and palmitates, are the chief components of toilet and laundry soaps. Other oleates, such as those of aluminum, copper, calcium, mercury, and zinc, are insoluble in water and are used as dry lubricants, paint driers, water repellents, dusting powders, and medicines. *See* CARBOXYLIC ACID; DRIER (PAINT); ESTER; SOAP AND DETERGENT. [E.H.H.]

Olefin sulfide

One of a number of organosulfur compounds,



also called episulfides or thiiranes, that are the sulfur analogs of olefin oxides. A general synthesis involves reactions of thiocyanate ion with olefin oxides. The episulfides are far less known than the epoxides, but have been extensively studied since 1950. They undergo some reactions which are analogous to those of epoxides, for example, ring opening with HCl, H_2S, RSH, and alcohols, ROH. The tendency of olefin sulfides to polymerize (to polysulfides) is quite marked, however, and frequently complicates their use in other reactions. *See* EPOXIDATION. [N.K.]

Oligocene

The third of the five major world-wide divisions (epochs) of the Tertiary Period (Cenozoic Era), the epoch of geologic time extending from the end of the Eocene to the beginning of the Miocene; the youngest epoch of the older Tertiary (Paleogene or Nummulitic). *See* CENOZOIC; TERTIARY.

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

Tertiary					Quaternary	
Paleocene	Eocene	Oligocene	Miocene	Pliocene	Pleistocene	Recent

In 1833 the British geologist Sir Charles Lyell subdivided the Tertiary into Pliocene (youngest), Miocene, and Eocene. Subsequently, it was realized that certain contemporaneous deposits were being classified as upper Eocene by some geologists and as lower Miocene by others. Accordingly, as a result of studies of these particular strata in Germany and Belgium, E. Beyrich in 1854 proposed the name Oligocene and indicated that it was intermediate between the older Eocene and the younger Miocene. The ensuing hundred years have witnessed the firm entrenchment of Oligocene as the middle, world-wide division of the Tertiary. Although the limits of the Oligocene are marked in places by physical breaks in the rock record, no physical change is present at other localities, probably the majority, and thus, differentiation of Oligocene from the Eocene rocks below and the Miocene ones above may be difficult. In such instances identification of Eocene, Oligocene, and Miocene is primarily a comparative paleontologic problem.

The Oligocene Series includes all rocks formed during the Oligocene Epoch, but the term is used most specifically with respect to the sedimentary rocks formed during this interval of the Tertiary Period. These latter contain the plant and animal remains which are the primary bases for the identification of Oligocene age.

Strata. The Oligocene strata include all the common sedimentary types, varying from marine through marginal-marine or intermediate to terrestrial in nature. They are typically unconsolidated to poorly consolidated and are widely dispersed throughout the world. The terrestrial beds are best known in the continental interiors while the marginal and marine beds are most widespread near the continental margins in the areas of the coastal plains and continental shelves. Especially noteworthy examples are present in (1) the Gulf Coastal province of the United States and Mexico; (2) the intermontane basins of the North American Cordillera; (3) the North Sea area of northern Europe, the Mediterranean Sea area of southern Europe and northern Africa, and the intracontinental basins of central and eastern Europe and southern Asia, (4) the Siwalik region of the Himalaya Mountains, and (5) the coastal region of South Australia. Most of the known marine and marginal Oligocene strata still are relatively flat-lying and occur near sea level, but crustal disturbances have appreciably deformed some to considerable elevations. Others have been depressed or have subsided below sea level. The terrestrial strata were deposited above sea level in the continental areas and, for the most part, have remained above this datum point. Ordinarily, they are somewhat more deformed than younger deposits and less modified than older ones. They contain important quantities of oil and gas, potable water, industrial rocks such as sand, clay, limestone, or marl, and other products.

Igneous rocks of Oligocene age include both intrusive and extrusive types, the latter being best known because they commonly occur as layered volcanic materials at the surface and are therefore more readily available for observation. Prominent among the Oligocene volcanic deposits are the thick, fossiliferous ash beds o the American Pacific Northwest.

the body wall and intestine. Sensory cells of various types are located in the epidermis and some members of the Naididae have eye spots. Oligochaetes react to a wide range of stimuli.

Reproduction. All oligochaetes are hermaphroditic. The testes are located anteriorly to the ovaries and both are derived from the septal peritoneum. The gametes are liberated into the coelom or pouches thereof and reach the outside through variously differentiated ducts. In copulation spermatozoa are received in spermatheca and at oviposition, which occurs later, the clitellum secretes a cocoon into which eggs and spermatozoa are discharged. Embryonic development occurs within the cocoon and is of the spiral determinate type. In the families Aeolosomatidae and Naididae, asexual reproduction by a type of binary fission (paratomy) occurs. *See* INVERTEBRATE EMBRYOLOGY; REPRODUCTION, ANIMAL.

Economic and theoretical importance. The oligochaetes have been used in studies of physiology, regeneration, and metabolic gradients. Some aquatic forms are important in studies of stream pollution as indicators of organic contamination. Earthworms are important in turning over the soil and reducing vegetable material into humus. It is likely that fertile soil furnishes a suitable habitat for earthworms, rather than being a result of their activity.

J. Stephenson and particularly W. Michaelsen have been outstanding students of the zoogeography of oligochaetes, but much remains to be done in this field. It is unlikely that the occurrence of certain families of earthworms in the Southern Hemisphere lends credence to the concept of continental drift. These families are more probably additional examples of southern relict faunas.

Classification. The Oligochaeta have been considered with the Polychaeta as an order of the Chaetopoda and with the [illegible]
the Clite[illegible]
and the [illegible]
chaetes [illegible] The Hirudinea are much closer to the oligochaetes, but constitute a homogeneous group long considered a separate class. The Oligochaeta, therefore, are here treated as a class of the phylum Annelida, coordinate in rank with the Polychaeta and Hirudinea as previously proposed by G. Pickford. Following Michaelsen, there are four orders of the class: (1) Plesiopora plesiotheca micronephridiostomal, male pores on the segment following the testes, and spermathecae in the region of the genital segments, (2) Plesiopora prosotheca, as for the first order, except the spermatheca are a number of segments in front of the genital segments, (3) Prosopora, mesonephridiostomal, male pores in the segment of the posterior testes, and (4) Opisthopora, meganephridiostomal, male pores opening posteriorly to last testicular segment.

That the oligochaetes are descended from marine polychaetelike ancestors seems certain, but there is no agreement as to the relationships of families within the class. Michaelson regarded the Aeolosomatidae as primitive; Stephenson, the Lumbriculidae, and no definite solution of this question has been reached. [P.C.H.]

Bibliography: W. Michaelsen, *Oligochaeta*, in W. Kükenthal and T. Krumbach, *Handbuch Der Zoologie*, vol. 2, 1928–1930; J. Stephenson, *The Oligochaeta*, 1930.

Oligoclase

A plagioclase feldspar with a composition ranging from $Ab_{90}An_{10}$ to $Ab_{70}An_{30}$ ($Ab = NaAlSi_3O_8$ and $An = CaAl_2Si_2O_8$). Natural material in the range from Ab_{98} to Ab_{83} is usually submicroscopically unmixed into domains of An_2 and $An_{25\text{-}30}$ composition. Oligoclases and albites that exhibit a blue luster as a consequence of this unmixing are called peristerite. If Fe_2O_3 is present as thin flakes oriented parallel to certain structurally defined planes, such oligoclase is called aventurine or sunstone. *See* FELDSPAR; GEM; IGNEOUS ROCKS. [F.L.A.]

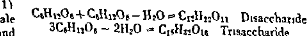
Oligomera

One of the three subphyla of the phylum Vermes proposed in 1910 by O. Bütschli. The Oligomera comprised those groups with two or three coelomic divisions. The subphylum was divided into four branches, the Tentaculata which contained the ectoproct Bryozoa and Phoronida, the Brachiopoda, the Chaetognaths, and the Branchiotremata or hemichordates. *See* AMERA; POLYMERA. [C.B.C.]

Oligosaccharide

A class of sugars which consists of a few monosaccharide units. Those sugars containing up to 6 units, many of which occur in nature, have been isolated as crystalline compounds. Fragments, obtained by controlled hydrolysis of various polysaccharides with acid, consisting of monosaccharides up to 10 units, are also termed oligosaccharides. *See* MONOSACCHARIDE.

Composition. The oligosaccharides may be considered as glycosides in which a hydroxyl (OH) group of one monosaccharide is condensed with the reducing group of another, with the loss of $n - 1$ molecules of water (n = number of monosaccharide residues). Thus



$$C_6H_{12}O_6 + C_6H_{12}O_6 - H_2O = C_{12}H_{22}O_{11} \quad \text{Disaccharide}$$
$$3C_6H_{12}O_6 - 2H_2O = C_{18}H_{32}O_{16} \quad \text{Trisaccharide}$$

If two sugar units are joined in this manner, a disaccharide results; a linear array of three monosaccharides thus joined by glycosidic bonds is a trisaccharide, and so forth. On the basis of the number of constituent monosaccharide units, the oligosaccharides are classified as disaccharides, trisaccharides, tetrasaccharides, and so on. No sharp distinction can be drawn between the oligosaccharides and polysaccharides, it is chiefly a matter of the latter's possessing higher molecular weights. *See* POLYSACCHARIDE.

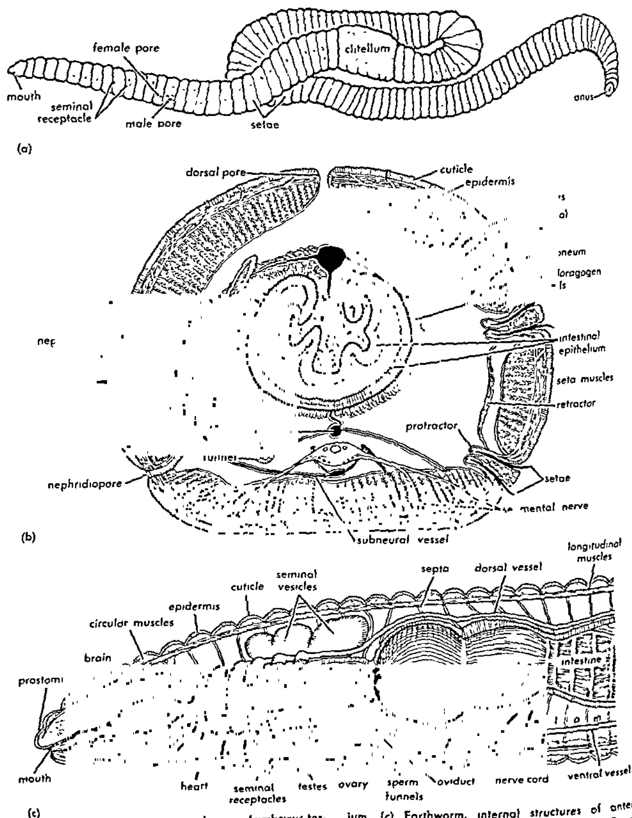


Fig. 2. Lumbricidae. (a) The earthworm, *Lumbricus terrestris*, external features (b) Earthworm, diagrammatic cross section. Left half shows an entire nephridium and a dorsal pore; right half includes setae but no nephridium. (c) Earthworm, internal structures of anterior portion from the left side (From T I Storer, General Zoology, 3d ed, McGraw-Hill, 1957)

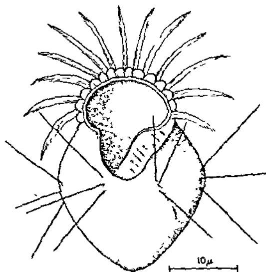
Naididae circulate water in and out of the posterior part of the intestine.

Circulation. The circulatory system consists of a main dorsal vessel, often arising from a perienteric sinus, in which the blood flows anteriorly, and various lateral, ventral, and subneural vessels The blood contains a respiratory pigment, erythrocruorin, and phagocytic corpuscles The contraction of the dorsal vessel or of lateral vessels, "hearts" propels the blood See RESPIRATORY PIGMENTS

Nervous system. The nervous system is composed of dorsal cerebral ganglia, circumesophageal connectives, and paired ventral nerve cords with segmentally arranged ganglia wh... off nerves

Oligotrichida

A minor order of the Spirotricha. If somatic ciliature is present, it is sparse. The bodies are round in cross section, and the adoral zone of membranelles is often highly developed at the anterior, or oral, end of the organism. Species are found in fresh and salt water habitats, and none occur as



Halteria, an example of an oligotrichid

parasites. *Halteria* (see illustration) has long bristles which are used in a kind of jumping movement. *Strombidium* is another frequently encountered species. See SPIROTRICHA. [J.O.C.]

Olive

The evergreen olive, *Olea europeae*, is among the most important of the subtropical fruit crops of the Mediterranean region (see EVERGREEN PLANTS). It is grown commercially in the United States only in California, where the annual production has a value of $9,000,000. A considerable number of the



Olive. (From L. H. Bailey, ed., The Standard Cyclopedia of Horticulture, vol. 2, Macmillan, 1937)

small white flowers borne in the spring frequently contain pollen only, hence are unable to produce fruit. Pollination is by wind. The fruit is a drupe of high oil content (40–65%) which is expressed by mechanical means. A bitter ingredient must be removed by soaking in lye before the fruit is edible (see FAT AND OIL, EDIBLE). Many varieties are grown for oil or processing or both. The queen olive refers to the large fruits of any variety used for food and eaten either green or black (ripe). See FRUIT (TREE). [C.A.S.]

Olivine

A name given to a group of magnesium-iron silicate minerals crystallizing in the orthorhombic system. Crystals are usually of simple habit, a combination of the dipyramid with prisms and pinacoids. The luster is vitreous and the color olive-green, giving rise to the name olivine. Hardness is 6½–7 on Mohs scale; specific gravity is 3.27–3.37, increasing with increase in iron content. See SILICATE MINERALS.

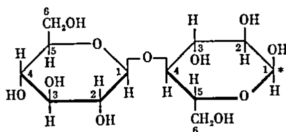
Olivine is a nesosilicate with composition $(Mg,Fe)_2SiO_4$. It comprises a complete solid solution series from the pure iron member fayalite, Fe_2SiO_4, to the pure magnesium member forsterite, Mg_2SiO_4. Minerals of intermediate composition have been given their own names but are usually designated simply as olivine. The magnesium-rich varieties are more common than those rich in iron. The minerals tephroite, Mn_2SiO_4, monticellite, $CaMgSiO_4$, and larsenite, $PbZnSiO_4$, although not in this chemical series, are sometimes included in the olivine group.

Olivine is found in some crystalline limestones but occurs chiefly as a rock-forming mineral in igneous rocks. It varies greatly in amount from an accessory to the main rock-forming constituent. Although it may be present in granites and other light-colored rocks, it is found chiefly in the dark rocks such as gabbro, basalt, and peridotite. The rock dunite is composed almost completely of olivine.

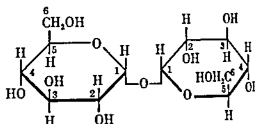
Olivine is one of the first minerals to form upon crystallization of a magma. It is believed that this early-formed olivine accumulated through the process of magmatic differentiation to form the large dunite masses. The type locality is at Dun Mountain, New Zealand; the rock is also found with corundum deposits in North Carolina. See MAGMA; PERIDOTITE.

Olivine alters readily to serpentine, a hydrous magnesium silicate. The alteration may take place on a large scale to form great masses of the rock serpentine, or on a small scale to form pseudomorphs of serpentine after single crystals of olivine. See SERPENTINE; SERPENTINITE.

At a few localities, notably on St. John's Island in the Red Sea and in Burma, olivine is found in transparent crystals. These are cut into gem stones which go under the name of peridot. Olivine is a major constituent of many stony meteorites (see METEORITE). [C.S.HU.]



Cellobiose (α form; *reducing group)
(reducing disaccharide)



Trehalose
(nonreducing disaccharide)

The oligosaccharides may be considered as glycosidic condensation products of the simple sugars, in which the second sugar unit serves as the aglycone group, that is the glycosidic hydroxyl of one of the constituent sugars is substituted in the same manner as is glucose in the α- and β-methylglucosides. If the union occurs in such a way that the reducing group of one of the sugars is left free, the complex sugar which is formed is reducing. It will mutarotate, form an osazone, and give the other carbonyl reactions of reducing monosaccharides. I[illegible]

ducing and will not give any of the reactions characteristic of a sugar with a free carbonyl group. The disaccharides, cellobiose and trehalose, both of which contain two D-glucopyranose residues, are examples of these two types of oligosaccharide. In the formula the asterisk denotes the reducing group

Most common disaccharides are dihexoses, although a few naturally occurring members of this group, such as primeverose, are known in which a pentose and a hexose are united together. The [illegible] be [illegible] two [illegible] ose

or raffinose. Sucrose consists of D-glucose and D-fructose, and raffinose consists of D-glucose, D-fructose, and D-galactose residues

With the exception of D-fructose, the various monosaccharide residues comprising the naturally occurring oligosaccharides have the pyranose structure, or a six-membered ring. When D-fructose serves as the glycosidic component in an oligosaccharide, it always occurs in the furanose form

Besides the many known naturally occurring free oligosaccharides, a great variety of this class of compounds can be obtained by enzymatic degradation, or by controlled hydrolysis of a polysaccharide with acid For example, the treatment of starch with amylases produces maltose. Under certain [illegible] cellobiose can be ob[illegible] ULOSE; MALTOSE. [illegible] iturally occurring [illegible] ablished common names, such as sucrose, lactose, melizitose, raffinose, stachyose, which were assigned before their complete structures were known. Rational names which indicate the chemical constitution of these

and the other known oligosaccharides have established jointly by the American and B[illegible] [illegible] nomenclature [illegible] as a gl[illegible] aldose (or glycosyl ketose) and a nonred[illegible] disaccharide as a glycosyl aldoside (or gl[illegible] ketoside) from its component parts Thus, t[illegible] ducing disaccharide α-lactose, consisting of lactopyranose united by a β-glycosyl linka[illegible] C-4 of D-glucopyranose is designated as 4-[illegible] galactopyranosyl-α-D-glucopyranose. A non[illegible] ing disaccharide, such as sucrose, which is [illegible] posed of D-glucopyranose and D-fructofu[illegible] united by α- and β-glycosyl linkages, is [illegible] α-D-glucopyranosyl-β-D-fructofuranoside, or [illegible] fructofuranosyl-α-D-glucopyranoside. A gl[illegible] of a reducing disaccharide, for example, me[illegible] lactoside, is designated as methyl-4-O-β D-g[illegible] pyranosyl-α-D-glucopyranoside. *See* LACTOSE; [illegible] FINOSE.

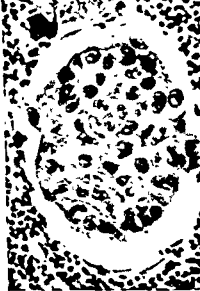
For naming oligosaccharides containing [illegible] than two units, the respective positions invol[illegible] the glycosidic linkages are indicated by two [illegible] bers and an arrow in parentheses. Thus, [illegible] ducing trisaccharide, maltotriose, is defined [illegible] α-D-glucopyranosyl-$(1 \rightarrow 4)$-O-α-D-glucopyra[illegible] $(1 \rightarrow 4)$-α-D-glucopyranose.

Properties. Oligosaccharides have the same [illegible] erties as their constituent monosaccharides, [illegible] as those properties may be modified by linki[illegible] units together. As an example disaccharide[illegible] alcoholic reactions just like the monosacch[illegible] but the number of reactive alcohol gro[illegible] smaller by 2 than the sum of the alcoholic [illegible] of the two monosaccharides. This is becau[illegible] hydroxyl position from each monosaccharid[illegible] is involved in the linkage between the two uni[illegible] constitute the disaccharide, and is not r[illegible] until after hydrolysis A hexose molecule ca[illegible] a pentaacetate by replacement of 5 hy[illegible] groups, but if 2 hexoses are joined to form [illegible] carbon disaccharide, 2 hydroxyl positions [illegible] pear by union, and only 8 of the original [illegible] available for replacement; consequently, th[illegible] acetylated disaccharide is an octaacetate. [illegible] larly, methylation takes place to the extent [illegible] troducing 8 methyl groups in a 12-carbon dis[illegible] ride molecule. [V[illegible]

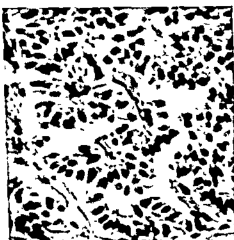
Bibliography: *Chem. Eng. News*, 31:1776, [illegible] *J. Chem. Soc* (London), pp [illegible] 1952.













(Top, left to right) Papanicolaou smear from the uterus Fibrosarcoma. Anaplastic carcinoma cells in a lymph node
(Bottom, left to right) Adenocarcinoma of the rectum Bronchial adenoma Epidermoid carcinoma of the skin
(Below, left to right) Adenomatous polyp of the stomach Adenocarcinoma of the rectum Liver metastases from an epidermoid carcinoma originating in the right lung



Omega

A long-distance continuous-wave navigation system of the hyperbolic type. This system utilizes frequencies in the 10–14 kc band. Emissions are from three or more stations, which are synchronized in phase. The transmissions from the stations are sequential and are followed by transmissions on a second frequency which is separated from the first by 500 cycles. All transmissions of the carrier frequency phase are stored by means of crystal oscillators. The phase differences between the 500-cycle beat notes are utilized to produce hyperbolic lines of position of relatively low accuracy. High accuracy is obtained by comparison of the phases of the carrier frequencies. *See* HYPERBOLIC NAVIGATION SYSTEM. [P.C.S.]

Oncology

The study of the causes, genesis, characteristics, classification, and treatment of tumors. Historically and etymologically the term tumor was applied to all types of tissue swellings regardless of their nature. In current usage the definition is limited to abnormal masses that are new growths of tissue discordant with the economy of the tissue of origin or the host's body as a whole. The word tumor is synonymous with neoplasm. Tumors consist of an accentuated proliferation of deranged living cells (parenchyma) and usually, but not invariably, a supporting connective tissue including blood vessels and lymph channels (stroma). They are a colony of permanently altered cells, and because they are derived from body cells there are at least as many types of tumors as there are types of body cells. The growth may be solid or cystic. On the basis of their form and pattern of growth, tumors are divided into two general groups: benign and malignant. *See* NEOPLASIA.

TUMORS

Benign tumors are usually slow-growing, localized, and circumscribed. Microscopically they have a relatively orderly architecture and the cells generally resemble adult body cells. In contrast, cancers are malignant tumors regardless of their appearance or tissue of origin. They tend to grow rapidly, to invade surrounding tissues, and to spread to distant sites in the body. Microscopically, cancer cells usually do not resemble adult cells and the tissue architectural patterns are often absent (see illustration). The smooth glistening lining of the rectum with orderly folds surrounds a cauliflowerlike mass of dull gray adenocarcinoma. The anus shows normal pigmentation. The rectum has been opened along its axis. When lesions such as the one shown are seen under a microscope, the orderly functioning glands of the lining of the rectum contrast sharply with the deep blue irregular cells of the cancer, which form glands having deranged size, shape, and position.

The details of the basic process whereby normal body cells become malignant are unknown. Many predisposing and contributing factors have [illegible] discovered.

Tumors are ubiquitous in mammals and [illegible]sent a problem equal to any in biological [illegible] and medical care. All extensively studied spec[illegible] higher animals have been found to develop tumors. In man, cancer is second only to heart and vascular diseases as a cause of death. About 450,000 new cases of cancer are diagnosed each year in the United States, and about 260,000 of these people will die of it. At this rate approximately one of every five persons living today will ultimately die of cancer. Though cancers may start at any age they occur with progressively increasing frequency in age groups beyond 40. In men the five most common sites in the body for cancer to occur are skin, lungs, prostate, stomach, and intestines. In women they are breast, uterus, intestines, skin, and stomach.

Characteristics. Benign and malignant tumors have some attributes in common. Both may begin at almost any site within the body. One cell or a small group of cells may begin to proliferate in an autonomous manner. Consequently, tumors are primarily composed of living cells and they are deranged descendants of body cells. Though more or less resembling the cell of origin, tumor cells usually vary from the normal in size, shape, and staining properties and have defective physical and spatial relationships with the surrounding tissue. Most significant is their accentuated proliferative vigor. This is a basic feature of all tumors, for without it, the prime characteristic, growth, would be lacking. Normally, cell proliferation is in equilibrium with cell loss. This meets the body's needs for maintenance and repair of tissues and organs. Tumor cells proliferate without regard for the demands of the body economy and often to the detriment of normal tissues. They also have a curious priority on nutrients within the body. A person dying of cancer may be severely malnourished yet the tumor tissue selectively and with great priority acquires the meager nutrients available and continues to grow. Some benign tumors behave similarly. Lipomas, benign fatty tumors, will continue to accumulate fat after the host has become starved and emaciated. Tumor growth tends to be progressive and without limitation, though long periods of quiescence may intervene between periods of rapid growth.

The physical form of benign and malignant tumors may sometimes be similar. When a tumor grows from a body surface such as the skin, the lining of the gastrointestinal tract, or uterus, it may become pedunculated in shape and have a thin stalk. These are called polyps. The illustration shows an adenomatous polyp of the stomach. Near the junction of the milky white esophageal lining and the stomach lining is a shiny, smooth, benign polypoid adenoma with a thin stalk.

The term polyp refers to physical shape rather than growth potentialities. Tumors may be solid or cystic If the cells are derived from secretory glandular tissue, abundant secretion may accumulate in the glandular spaces. This results in a cystic appearance. Pressure within a tumor may strangulate some blood vessels, causing the death of some zones This may also produce a cystlike appearance A tumor with cystic change should not be confused with nonneoplastic simple cysts. The

from normal tissue primarily in general architecture and excess of cells. Bronchial adenoma is shown in the illustration. Microscopically, orderly chains of cells are lined up like pickets in a fence *in this benign tumor*. It has a capsule of connective *tissue*. *This is* in contrast with malignant glandular tumors (adenocarcinomata) The slowly expanding mass of cells compresses the surrounding tissues The growth tending to be *centripetal* and cohesive is therefore well demarcated from the surrounding tissues Frequently compression and overgrowth of the adjacent connective tissue produces a capsule Benign tumors are dependent on the *stroma* of the site *of origin* and do not invade the surrounding tissues or spread to distant sites Growing progressively and unrestrained by body regulatory mechanisms, the mass may attain a very large size Tumors *weighing* more than 50 lb have been recorded

Death is rarely caused by benign tumors. If inadequately excised they may recur If not excised they may achieve a size which interferes with the function of an organ and results in a *health hazard* For example, a benign *adenoma in* the bronchus may achieve sufficient size to obstruct the flow of air to the lungs *A uterine leiomyoma* (fibroid) may be so situated as to cause hemorrhage

Malignant tumors (cancers), on the other hand, have in addition to some similarities with benign tumors many unique features The three cardinal properties of malignancy are anaplasia, invasion, and metastasis These properties refer to the appearance and behavior of cancer cells in which they usually differ greatly from benign tumor cells (see illustration). They reflect the independence of cancer cells, the ability to grow at sites other than the locus of origin

Anaplasia Anaplasia is the structural aberration of cancer cells which results in a more primitive (embryonic) appearance and diminished or absent adult function Their size, shape, staining properties, and spatial relationships to one another are usually markedly deranged as shown in the illustration of anaplastic carcinoma cells which are seen in a lymph node They are relatively large, irregular *in* shape, with large *irregular nuclei* stained deep *blue* These anaplastic cells bear no resemblance to the *breast* tissue *from* which they were derived. Anaplastic cancer cells bear scant resemblance to adult forms. Generally with greater degrees of anaplasia the growth rate increases. However, *notable exceptions* exist. Anaplastic cancer cells lack orientation with respect to the tissue as a whole. Instead of an orderly spatial arrangement their distribution is often jumbled.

Invasion. This phenomenon is a distinguishing feature of cancers Rather than remaining circumscribed, cancer cells tend to invade and infiltrate surrounding tissues, often destroying them as exemplified by the illustration of an epidermoid carcinoma of the skin Microscopically, beyond the edge of the *spreading primary lesion the individual* carcinoma cells *are seen invading the connective* tissue. Like the root of a plant *extending* into soil, cords of cancer cells grow into surrounding tissues. They follow lines of least resistance Where surrounded by a tissue of uniform consistency, cancer cells spread equally in all directions How they accomplish invasion remains a *mystery*

Metastasis The establishment of secondary sites of cancerous growths at a *distance from* the original site is metastasis. The *illustration* shows liver metastases from an epidermoid *carcinoma* originating in the right lung. Large gray *cannonball*-like masses of carcinomatous *tissue are scattered throughout the liver*

When cancer cells invade veins or lymph channels (rarely body cavities or surfaces), they circulate *to distant* parts of the body where they may *become attached*, attract a stroma at the new site, and begin to proliferate. After a *cancer has* metastasized excision of the *primary site has* no significant effect on the progression of the *metastases*. To be effectively cured cancers must be *removed* early in their development before metastases have occurred.

Some cancers appear to be *independent* from their *inception, and able* to invade and metastasize *immediately* Others *undergo* a *series* of changes *in form* before becoming *malignant*. Precancerous change is the stepwise modification of body cells toward a *cancerous* appearance but *before* becoming cancerous Sometimes this may take *many* years Cancer of the uterine cervix is an example. Repeated studies of these cells reveal that they may resemble cancer for 8–10 years before actually *becoming cancer with* consequent invasion and metastasis In some patients who have precancerous changes the evolution to cancer may never occur Progression is not inevitable. Converselv the stepwise chan...
Even when
the precanc
of invasion
...metastasis can be extremely slow at first Only later when these *cancer* cells have undergone further evolution do *they invade* and metastasize with great rapidity and lead to a fatal outcome Some cancers *may* seed down metastases which remain quiescent for many years after surgical removal of the primary cancer, then begin to

Basal cell carcinoma One of the most common tumors, the so-called basal cell carcinoma, is not included in the table because of the difficulty of classification. It begins as a small nodule in the skin which ulcerates, has a ragged, dirty, irregular appearance, and gradually increases in size. The tumor remains local, rarely if ever metastasizing, but will invade, compress, and destroy surrounding tissues It erodes through cartilage and bone, hence the term rodent ulcer is sometimes applied Therefore, the invasive properties of the tumor are consistent with malignancy However, its failure to metastasize suggests a benign nature. As with other benign tumors, complete local excision effects a complete cure.

In contrast, epidermoid carcinoma is a truly malignant skin tumor It may arise wherever the stratified squamous type of epithelial cells may be found, such as skin mouth, esophagus, bronchus, and uterine cervix. Beginning as an elevated, papillary, hard nodule in the epithelium, it more or less rapidly infiltrates the deeper connective and adipose tissues spreading to the regional lymph nodes. If the primary tumor is not treated, metastases may occur to internal organs, principally lungs or liver.

Adenocarcinoma Adenocarcinomas may arise wherever glandular tissue is found They frequently arise in the breast, rectum, colon, stomach pancreas, prostate, and uterus Breast adenocarcinoma, derived from the lining of breast ducts, begins as a small, single, firm nodule which has an irregular border It can be moved about only with difficulty and to a limited degree. Sometimes the nipple or overlying skin will be drawn inward or it will have an orange peel appearance A very few will be painful

Microscopically, breast cancers have a wide variety of forms, some of which spread and metastasize with great rapidity, whereas others are much more sluggish They first metastasize to the lymph nodes in the armpit and chest wall Later they may spread to lungs, liver, or bones Unlike breast carcinomas, which are often noticeable when they are small adenocarcinomas of the colon or rectum may achieve considerable size before being discovered They tend to infiltrate and metastasize with vigor At the other extreme adenocarcinomas of the prostate are exceedingly common in elderly men, but they uncommonly cause symptoms or spread outside the prostate gland

Autopsy studies reveal that 20% of the men over 40 years of age have cells resembling latent adenocarcinomas of the prostate gland A small percentage of these will evolve further into a more rapidly growing infiltrating metastasizing carcinoma that produces urinary tract symptoms Rarely in younger men the latter circumstance may prevail from the tumor's inception. Neuroblastoma, a malignant tumor derived from nervous system cells, is the most common cancer of infants Occasionally the tumor is present at birth

Lymphoma-leukemia tumors. The lymphoma-leukemia group of tumors has many unique features Delineation into benign and malignant forms is difficult. Lymphoid tissues are specialized descendants from the cells of embryonic mesenchyme (primitive connective tissue). These cells normally undergo a series of changes from mesenchymal cells to reticulum cells to lymphoblasts to mature lymphocytes. Aggregates of these cells are in the spleen and scattered throughout the body in small nodules called lymph nodes Mature lymphocytes circulate in the blood as one type of white blood cell assisting in the body's defense against infections Lymphomas are tumors that arise as defective descendants of any of the above cell types, causing enlargement of the lymph nodes and spleen If the cells circulate in the blood in large numbers, the process is called lymphatic leukemia. Similar mesenchymal cells located in bone marrow undergo a series of changes leading to the formation of other types of white blood cells (myelocytes), red blood cells, and platelets. Leukemias composed of these white blood cells are called myelogenous, whereas those from the lymph nodes are called lymphogenous or lymphoid. Those that proceed rapidly to a fatal termination are called acute. Leukemias are the second most common neoplastic disease of children The chances of cure are poor Death often results from infections or hemorrhage. The defective lymphocytes lack the adult functions that make them protectors against infection Secondary disturbances in the clotting mechanism may lead to fatal hemorrhage In elderly people the leukemia is often chronic and without fatal termination.

ETIOLOGY

Since tumors are of diverse types derived from virtually [illegible]
variety of [illegible]
to have [illegible] basically, the change from a normal cell to a neoplastic one involves a change in cellular heredity. The tumor daughter cells inherit their altered form and behavioral pattern from their parent cell and pass the alteration to subsequent cells Thus, malignant cells are derived from previously normal ones Etiology is the study of the causes of this transformation The change from normal to cancer may be direct or it may involve a stepwise series of changes The interaction of an appropriate etiologic agent with susceptible tissue cells produces the neoplastic changes in form The altered behavioral activity of the cells occurs concomitantly Carcinogenesis is the term applied to this process of change. How cells that normally divide their energy between proliferation (growth) and specialized function (work) subordinate the [illegible]

[illegible] tumors has been conducted since 1930. In 195[illegible] more than 1000 research projects attacking variou[illegible]

grow. Indeed, the rate and pattern of growth of each cancer differs from all others.

Pseudotumor. Pseudotumor is a term sometimes applied to nonneoplastic masses of diverse types and causes. They differ from benign and malignant tumors because they are self-limiting and are the body's attempt to reestablish equilibrium in response to injury. Tissue swellings as a result of inflammation are common. Blood vessel congestion, accumulation of body fluids, and reparative cells produce a mass at the site of inflammation which disappears when the offending agent is overcome. If a circumscribed area of tissue death occurs, an abscess may persist. This too represents a nonneoplastic mass. In the reparative process an orderly overgrowth of connective tissue in excess of adequacy may ensue in some individuals, producing an elevated, hard, scar tissue mass called a keloid. Negroes are particularly prone to develop keloids. The orderly overgrowth of tissues to meet an unusual body demand is termed hyperplasia. The growth of a skin callus on a laborer's hand is an example. Excessive skin growths in response to certain infectious agents are condylomas. Simple cysts are another type of pseudotumor. They are accumulations of nonviable material, often fluid, within a cavity surrounded by a definite wall. This often occurs as a result of obstruction of a glandular duct preventing the egress of secretions. See HYPERPLASIA.

Inflammation may secondarily add further to the size of the mass. Sebaceous cysts (wens) belong to this group.

Classification. Because of their widely differing growth patterns, each type of benign and malignant tumor has a different implication of illness and life expectancy. It is natural, therefore, that oncologists expend considerable effort attempting to classify them systematically. The aim is to group tumors of like form and behavior together in each category. Unfortunately, attempts at systematic classification are only partially successful for two major reasons. First, tumors of innumerable types occur; the gradations in appearance of each type merge one with the other so that establishment of categories must often be arbitrary. Second, to observe the static form of cells microscopically, then to infer growth behavior, is fraught with inaccuracy. Though anaplasia is often correlated with malignancy, numerous notable exceptions exist. Thus many tumors defy classification, for they have features in common with several categories. See ANAPLASIA.

valuable, for therein lies the hope for cancer prevention. Unfortunately so little is known about the genesis of cancer that this method of classification cannot be used.

Carcinoma. The table is a commonly used list of tumors based on the tissue of origin, cell type, and the presence or absence of malignant features. Those tumors derived from the internal and external body surface coverings and their derivatives such as the skin, sweat glands, breast, lining of the respiratory, gastrointestinal, urinary, and genital systems are epithelial tumors. Malignant tumors derived therefrom are called carcinomas. When a carcinoma takes origin from the skin or comparable epithelium, it is called epidermoid carcinoma. Benign skin tumors are called papillomas. When malignancies are derived from glandular epithelium, they are called adenocarcinomas. Adenomas are benign glandular tumors. For example a benign tumor of bronchial glands is called a bronchial adenoma. A malignancy derived from these glands is called bronchial adenocarcinoma. Though remotely descended from embryonal ectoderm, tumors of the nervous system have many features peculiar to themselves and represent a separate subgroup. For the same reason tumors derived from the placental epithelium also form a separate subgroup.

Sarcoma. The term sarcoma is applied to malignancies arising from the derivatives of embryonal mesoderm such as connective tissues, bone, muscle, cartilage, blood vessels, and blood cells. A benign fibrous connective tissue tumor is a fibroma whereas a malignant one is a fibrosarcoma. The illustration shows a fibrosarcoma in which microscopic streams of spindle-shaped cells are intertwined. Their appearance is in contrast with the carcinoma cells of the various carcinomata. A major subgroup is the tumors composed of blood cells, the lymphomas and leukemias. Mixed tumors and teratomas comprise a third major category of tumors. They are composed of mixtures of cells of the above categories. In all three categories the Greek word root "oma," meaning tumor, is applied to the end of the name of the involved tissue cell type.

The table lists only a few representative tumors in each category.

Tumor classification

Tissue of origin	Benign tumor	Malignant tumor
Epithelial tissues		
Skin	Papilloma	Epidermoid carcinoma
Glands	Adenoma	Adenocarcinoma
Nervous tissue		
Nerve cells	Neuroma	Neuroblastoma
Supportive cells	Glioma	Glioblastoma
Pigment cells	Melanoma	Melanocarcinoma
Placenta	Hydatidiform mole	Chorionic carcinoma
Connective tissues		
Adult	Fibroma	Fibrosarcoma
Embryonic	Myxoma	Myxosarcoma
Cartilage	Chondroma	Chondrosarcoma
Bone	Osteoma	Osteogenic sarcoma
Fat	Lipoma	Liposarcoma
Smooth muscle	Leiomyoma	Leiomyosarcoma
Lymphoid tissue	Lymphoma	Lymphosarcoma
Bone marrow		Leukemia
Mixed tissues		
Ovary or testis	Teratoma	Teratocarcinoma
Salivary gland	Mixed tumor	[illegible]d tumor

Basal cell carcinoma. One of the most common tumors, the so-called basal cell carcinoma, is not included in the table because of the difficulty of classification It begins as a small nodule in the skin which ulcerates, has a ragged, dirty, irregular appearance, and gradually increases in size The tumor remains local, rarely if ever metastasizing, but it will invade, compress, and destroy surrounding tissues It erodes through cartilage and bone; hence the term rodent ulcer is sometimes applied Therefore, the invasive properties of the tumor are consistent with malignancy. However, its failure to metastasize suggests a benign nature As with other benign tumors, complete local excision effects a complete cure

In contrast, epidermoid carcinoma is a truly malignant skin tumor. It may arise wherever the stratified squamous type of epithelial cells may be found, such as skin, mouth, esophagus, bronchus, and uterine cervix Beginning as an elevated, papillary, hard nodule in the epithelium, it more or less rapidly infiltrates the deeper connective and adipose tissues, spreading to the regional lymph nodes If the primary tumor is not treated, metastases may occur to internal organs, principally lungs or liver

Adenocarcinoma. Adenocarcinomas may arise wherever glandular tissue is found They frequently arise in the breast, rectum, colon stomach, pancreas prostate and ...

... border it can be moved about only with difficulty and to a limited degree Sometimes the nipple or overlying skin will be drawn inward or it will have an orange-peel appearance A very few will be painful.

Microscopically, breast cancers have a wide variety of forms, some of which spread and metastasize with great rapidity, whereas others are much more sluggish They first metastasize to the lymph nodes in the armpit and chest wall Later they may spread to lungs, liver or bones Unlike breast ... tice- ab ... the ro ... size be ... undiscovered They tend to infiltrate and metastasize with vigor At the other extreme, adenocarcinomas of the prostate are exceedingly common in elderly men but they uncommonly cause symptoms or spread outside the prostate gland

Autopsy studies reveal that 20% of the men over 40 years of age have cells resembling latent adenocarcinomas of the prostate gland A small percentage of these will evolve further into a more rapidly growing infiltrating metastasizing carcinoma that produces urinary tract symptoms. Rarely in younger men the latter circumstance may prevail from the tumor's ... malignant is the most the tumor ... carcinom.

Lymphoma-leukemia tumors. The lymphoma-leukemia group of tumors has many unique features. Delineation into benign and malignant forms is difficult. Lymphoid tissues are specialized descendants from the cells of embryonic mesenchyme (primitive connective tissue). These cells normally undergo a series of changes from mesenchymal cells to reticulum cells to lymphoblasts to mature lymphocytes Aggregates of these cells are in the spleen and scattered throughout the body in small nodules called lymph nodes Mature lymphocytes circulate in the blood as one type of white blood cell assisting in the body's defense against infections. Lymphomas are tumors that arise as defective descendants of any of the above cell types, causing enlargement of the lymph nodes and spleen If the cells circulate in the blood in large numbers, the process is called lymphatic leukemia. Similar mesenchymal cells located in bone marrow undergo a series of changes leading to the formation of other types of white blood cells (myelocytes), red blood cells, and platelets. Leukemias composed of these white blood cells are called myelogenous, whereas those from the lymph nodes are called lymphogenous or lymphoid. Those that proceed rapidly to a fatal termination are called acute Leukemias are the second most common neoplastic disease of children The chances of cure are poor Death often results from infections or hemorrhage The defective lymphocytes lack the adult functions that make them protectors against infection Secondary disturbances in the clotting mechanism may lead to fatal hemorrhage. In elderly people the leukemia is often chronic and without fatal termination

ETIOLOGY

Since tumors are of diverse types derived from virtually every tissue in the body and have a wide variety of growth patterns, they might be expected to have more than one cause. Basically, the change from a normal cell to a neoplastic one involves a change in cellular heredity The tumor daughter cells inherit their altered form and behavioral pattern from their parent cell and pass the alteration to subsequent cells. Thus, malignant cells are derived from previously normal ones. Etiology is the study of the causes of this transformation The change from normal to cancer may be direct or it may involve a stepwise series of changes The interaction of an appropriate etiologic agent with susceptible tissue cells produces the neoplastic changes in form. The altered behavioral activity of the cells occurs concomitantly. Carcinogenesis is the term applied to this process of change How cells that normally divide their energy between proliferation (growth) and specialized function (work) subordinate the latter ... the former ...

... eti ... behavior of tumors has been conducted since 1930 In 1953, more than 1000 research projects attacking various

aspects of the cancer problem were in progress in the United States. If a better understanding of the causes and mechanism of genesis of tumors were achieved, the means of prevention could then be devised.

Additional knowledge of the biologic behavior of tumors (the interaction of the tumor tissue with the host, especially invasion, metastasis, and host resistance) could also form the basis for development of therapeutic methods for those who have neoplastic diseases. This latter area of research attempts to find what the change in the relationship between normal and neoplastic cells is and how cancers infiltrate and metastasize.

Three techniques of experimentation have facilitated research on the biologic behavior of tumors. Tissue culture enables scientists to grow bits of tumor tissue in nutrient artificial media in glass containers. This enables direct observation of cell behavior in carefully controlled conditions for measurement. Secondly, tumor transplantation from one animal to another is a useful tool in the study of the host responses to the tumor, and the tumor's response to different hosts. Transplantation experiments have shown that the ability to metastasize and invade closely parallels the ability to attract or induce a stroma. The third method of studying the biological behavior of tumors is the application of immunologic techniques. It attempts to measure the host's development of immunity to tumor cell growth as a means of resistance The immune mechanism is presumed to be akin to that which occurs with bacterial or virus infections.

FACTORS IN CARCINOGENESIS

The factors that contribute to or encourage the neoplastic change may be divided into intrinsic (internal or originating within the body) and extrinsic (external or environmental) for convenience of description More than 400 carcinogenic agents are known, and others are suspected. These involve virtually every aspect of the human environment. The agents range from physical (radiation, trauma) to chemical, genetic, bacterial, viral, and ontogenetic (embryologic development of the individual).

Intrinsic factors. The genetic concept of the origin of cancer is one of the oldest intrinsic factors to be considered Nineteenth-century scientists noted aberrant distribution and shape of chromosomes during some cell divisions. They reasoned that such cells under appropriate conditions may lead to the formation of cancer. In an application of the genetic concept, many strains of mice have been selectively inbred so that they spontaneously develop cancers or are susceptible hosts to which human cancers can be transplanted It has not been proved that the findings in these animal experiments apply to humans Genetic makeup may affect susceptibility or predisposition, but the stimulus for carcinogenesis is usually not genetic. Studies of human family trees have contributed little support for the genetic concept. Cancers are not inherited, although some slight predisposition may be inherited. With the knowledge that one in five people will die of cancer, it is quite understandable that on the basis of chance alone, some families may have several members so afflicted. This does not necessarily mean that cancer runs in the family.

Hormones have been shown to play a role in the growth of some cancers and represent a second intrinsic factor. Some inbred strains of mice will not develop carcinoma of the breast unless large doses of estrogenic hormones are administered. This relationship has not been demonstrated in humans. However, the continued growth of some human breast adenocarcinomas is inhibited by removal of the ovaries, the site of estrogen production. Unfortunately, this response is often temporary. In men, removal of the testes, site of male hormone production, often decreases the rate of growth of some prostate adenocarcinomas and results in remissions or occasionally arrests the disease Therefore, some tumors are dependent on hormones, at least to a moderate degree

Extrinsic factors. Extrinsic stimuli associated with the onset of cancer are numerous. Whether each is the sole carcinogenic stimulus for a specific type of cancer or whether various stimuli may be additive is not known. Special types of viruses have been shown to cause certain neoplasms in animals Shope papilloma of the skin of rabbits, leukemia in chickens, and some breast carcinomas in mice are examples of virus-induced animal tumors. Viruses have not been discovered in human cancers in spite of intensive search. *See* TUMOR VIRUSES

Radiation An extrinsic carcinogenic agent of considerable public interest since the discovery of atomic fission is radiation. The types of high-energy electromagnetic radiations of carcinogenic interest are x-rays, gamma rays, electrons, neutrons, and protons The frequency of radiation-induced cancers in humans is difficult to assess because accurate information on exposure is usually lacking. Such factors as measured dosage of radiation, whether exposure is intense for a short period or light for a long period, volume of tissue exposed, and type of exposure are usually not accurately known. Thus, most human cancers associated with radiation are poorly documented.

One of the first recorded cases of radiation-induced cancer occurred in Hamburg, Germany, in 1902. A young man employed in an x-ray tube factory tested the tubes by repeatedly x-raying his own hand An extensive skin rash developed on his hand and, after 3 years, a cancer formed in the skin of his hand It metastasized to other regions of his body Subsequently, cases of radiation-induced cancers of bones, lungs, skin, and blood-forming tissues have been described Most notable are the radium-dial painters. In the manufacture of luminous watch dials luminescent radioactive material was used Some painters pointed their brushes by wetting them with their li[illegible] bone tu

osteogenic sarcoma) subsequently developed ny of the exposed painters. Under proper ions exposure to the luminescent material is ess because the dosage is infinitesimal.

kemias can be induced in man by radiation, ie incidence in exposed people is very low. cases of leukemia occur spontaneously without radiation The incidence of leukemias varies ferably with respect to race, social class, and aphic location Thus, a group of people who en to be exposed to radiation may have a r incidence of leukemias than the general lation for reasons other than their exposure such group are radiologists. The survivors of tom bombing of Hiroshima and Nagasaki have an increased leukemia rate The 194 000 exd survivors had 95 cases of leukemia by 1955 7 per 100 000 per year rate compares with the per 100,000 in the United States However, if selects the population from certain exposed s the rate increases to 23 per 100,000 per year statistics are strongly suggestive, but not enly conclusive, that a correlation exists Coned observation of the exposed individuals may vide a definitive answer in the future Many es of cancers have been experimentally induced nice, rats, guinea pigs, and rabbits by radiation re significant than the cancers produced is the t that so *few of the total* number of animals posed *actually* develop cancer Sunshine may o be a carcinogen Light complexioned peoples e a higher incidence of skin cancer on the exsed parts of the body such as *the face neck or* nds than do the *pigmented races* This is freently seen in fair-skinned people who work in the rect *intense* sunlight such as white sailors and rmers in the tropics Whether cosmic rays have n effect or not is completely unknown *See* RADIAION INJURY (BIOLOGY), ULTRAVIOLET RADIATION BIOLOGY).

Chemical carcinogens. Carcinogenic chemicals re numerous In 1775, Sir Percival Potts observed hat cancers of the scrotum occurred frequently in himney sweeps The chemical substance involved was isolated 140 years later by painting coal tars on the ears of rabbits Carcinomas were produced Subsequently numerous chemicals of the benzanthracene group have been identified as carcinogens Persons excessively exposed to aniline (which is excreted in the *urine*) have 33 times more cancer of the urinary bladder than the general population A partial list of the hundreds of chemicals that are carcinogenic in appropriate dosage includes beta naphthylamine, benzidine, some vital dyes, chromates, arsenic, nickel, and asbestos

Food None of the standard foodstuffs natural or purified have been shown to be carcinogenic Because cancers derive their energy from the host, it might be expected that the nutritional status of the host may in some way affect the neoplastic process Obese people have a higher over-all cancer rate than the general population. On the other hand, the high rate of liver carcinoma in some primitive tribes, notably the Bantu in Africa, has been ascribed to malnutrition

Air pollution. Atmospheric pollution by industrial soot and vehicular exhaust has become a problem of increasing concern since the advent of internal combustion engines. Although clearly proved cases of carcinoma caused by atmospheric pollution have not been found, population statistics show the incidence of lung carcinoma to be higher in urban than rural environments. Numerous carcinogens have been identified in polluted air, however, the dosage and conditions of exposure are poorly understood Obviously, urban and rural living differ in many more respects than merely atmospheric pollution. Therefore, interpretation of crude population statistics *must be cautious. See* AIR POLLUTION CONTROL.

Tobacco. Tobacco is another potential source of chemical carcinogens Since 1900 the deaths ascribed to lung carcinoma have increased. It is now the second most common type of cancer in men. Statistically the increase has *paralleled the increase* in cigarette smoking. Whether these two facts are related or not is disputed Much of the increased lung cancer incidence can be ascribed to better case finding (*chest x-ray, periodic physical exams, cancer publicity*) *and accurate* recording (tissue biopsy diagnosis, postmortem exams, tumor registry statistics) The average adult inhales more than 20 000 liters of *air per day* *Of the* many carcinogenic substances *contained* in air and dust to which man's lungs are exposed daily, tobacco smoke may be one of many factors which act in concert. The degree of importance of each factor remains controversial

Irritation and infection Chronic irritation and chronic infections are agents which occasionally are associated with the onset of cancer The site in the urinary bladder where the flatworm parasite *Schistosoma hematobium* infests is often the site of cancer formation (see SCHISTOSOMIASIS). Lung cancers occasionally originate in the walls of old tuberculous lesions In Kashmir, people who keep warm by carrying *a kangri*, a wicker basket containing an earthenware pot filled with smoldering leaves, against their abdominal skin frequently develop skin cancer at this site. Though continuous or intermittent irritation for many years has been associated with the onset of cancer, there is virtually no evidence that single physical trauma such as a bruise or contusion brings about cancerous change The reported cases, after careful documentation, prove that the trauma merely called attention to a preexisting cancer.

DIAGNOSIS AND TREATMENT

The most important tumor diagnostic tool is the oldest and simplest, a careful physical examination Tumors being space-occupying masses, they are frequently detected by direct vision or palpation Cancers of the skin and female breast are among the most frequently occurring types. These sites are ideal for early diagnosis if appropriate

attention is directed to them. Enlightened self-interest dictates that careful self-examination be done as a means to early diagnosis. Physicians often enhance their area of direct visibility by endoscopy. This is the insertion of an optical instrument into body orifices (esophagus, trachea, ears, or urinary, genital, or anal passages) to render visible lesions at these more remote sites.

Diagnostic x-rays are another means of visualizing tumors. Blood chemistry studies and blood counts may at times reveal clues of the existence of a cancerous process. Because noncancerous lesions such as inflammations and degenerations may visibly resemble cancer, the diagnosis of cancer is not proved until a biopsy is done. A biopsy is the removal of a small bit of tissue from a patient by a surgeon who then submits it to a pathologist for examination. By appropriate processing and staining, a microscope slide is prepared from the tissue. The pathologist observes the characteristics of the tissue under a microscope and informs the surgeon of his findings.

Since 1945, exfoliative cytology has gained extensive use as an aid to the early diagnosis of cancer. Cells that flake off from various body surfaces are stained and observed under the microscope An example is a Papanicolaou smear from the uterus observed microscopically, in which the exfoliated normal cells surround several epidermoid carcinoma cells. The latter cells are variable in size, shape, and dark staining, and have large irregular nuclei. This smear was made from secretions during a routine physical examination. This technique has been most effective in the early diagnosis of carcinoma of the cervix of the uterus In this way precancerous lesions can also be found and treated Therefore, a combination of intelligent self-examination, thorough medical examination, and exfoliative cytology would enable the early diagnosis of most of the common types of cancer

The three major types of weapons for the treatment of cancer available to physicians are surgery, radiation, and chemotherapy. Complete surgical excision of a benign tumor is usually sufficient to produce a cure. Malignant tumors can also be cured if they have not infiltrated adjacent tissue or metastasized by the time of excision. Radiation therapy, including x-ray and radioactive metals (cobalt, radium, phosphorus, iodine, and others), has application in the treatment of many types of cancers. Strong beams of radiation are focused on the tumor causing the death of susceptible cancer cells and scar formation. Thus, radiation which may produce cancer at one dosage level can destroy it at another. Whether radiation or surgery or both are used in the treatment of a particular neoplasm depends on many variables including the exact type of neoplasm, its size, location, rate of spread, and general condition of the patient. Different cancers differ widely in their susceptibility to destruction by radiation. Some are radioresistant. Others are radiosensitive In general, the more anaplastic tumor cells are radiosensitive, and the more mature types of tumors are radioresistant. Numerous exceptions to this rule exist. *See* RADIOLOGY.

Chemical agents that selectively destroy cancer cells and leave the body cells relatively unharmed have been developed. Chemical derivatives of mustard gas were among the first cancer chemotherapy agents. They were effective in retarding the growth of several types of tumors. Because it has been recognized that the growth of some tumors is hormone-dependent, chemical derivatives of hormones have been synthesized for their antitumor properties. These have on occasion restrained the growth of cancer, at least temporarily. *See* CHEMOTHERAPY

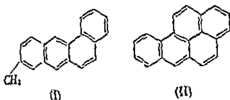
The most important aid to effective cure is early diagnosis. An estimated 25% more cures could be achieved if the diagnostic and therapeutic methods known at present were used optimally [N K M]

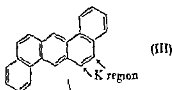
BIOCHEMISTRY OF CARCINOGENESIS

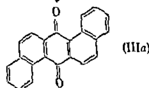
Carcinogenesis may be defined as the induction of neoplastic change in animal or plant tissues by some agent. Carcinogenic agents comprise a variety of chemical compounds, certain viruses, and physical agents such as ionizing radiations Chemical carcinogens include a variety of chemical groupings. The best known are some polycyclic and heterocyclic hydrocarbons, some azo dyes, certain aromatic amines, some carbamates, and others. A series of plastics that give rise to malignant tumors under very limited conditions of test are difficult to classify in this context as being either chemical carcinogens or physical mediators of this effect

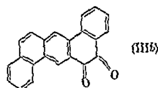
The biochemistry of carcinogenesis is a divers subject concerned with the metabolism of the individual agents Because knowledge of the biochemical difference between normal and neoplastic tissue is incomplete, no unified approach toward a clea goal is yet possible

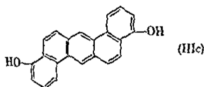
Biological response. The biological response t these agents is complex and depends upon the precise conditions of the test system. Potent compounds of the polycyclic hydrocarbon groups, such as 20-methylcholanthrene (I) and 3,4-benzpyren (II), give rise to skin cancer (squamous cell carci noma) in many species when applied to the skin i relatively low dosage, they induce subcutaneou malignancies (sarcomas) when injected in the sub cutaneous tissues. As little as 0 4 microgram o 1,2,5,6-dibenzanthracene (III) is an active do in the mouse. Similarly these agents will give ris to malignant tumors in many other tissues wit which they may be placed in contact. These compounds also manifest remote effects, these are mos easily observed in special strains of animals Lun adenomas occur following administration by an route of these agents in Strain A mice; leukemia i enhanced in the Ak and other strains. The azo dyes on the other hand, act largely when taken orally inducing hepatoma in rats, they act in other species, too, but less potently. The aromatic amine are best known for their induc[illegible] bladder can cer in the dog. Urethane, [illegible] ght

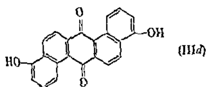


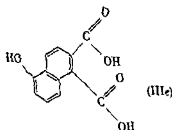


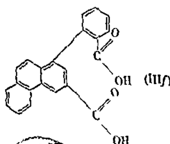












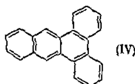
rise only to lung tumors (adenomas) is now known to be a versatile carcinogen, initiating skin carcinogenesis and also giving rise to mammary tumors and hemangioendotheliomas in mice.

In a variety of studies of skin carcinogenesis in mice and rabbits, it has been found that a biological mechanism involving at least two stages may apply. Thus carcinogenesis may be initiated by a small, subeffective dose of a carcinogen, which gives rise to morphologically undetectable latent tumor cells that may, even after many months, be converted into actual tumors by a promoting agent such as croton oil that need not be one of the carcinogens.

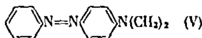
Investigations. Many pathways of chemical degradation of various chemical carcinogens have been elucidated However the investigator is still confronted with the major problem of differentiating between those changes that may be pertinent and those that may be quite irrelevant Several of the polycyclic hydrocarbons have been well studied and follow similar pathways. Studies with 1,2,5,6-dibenzanthracene may be considered to be representative of this class of compound; hydroxy derivatives, some quinones, and a phenolic acid (*IIIa-e*) have been isolated from urine, feces, and liver in various in vivo studies, indicating the diversity of changes undergone In addition, a dicarboxylic acid (*IIIf*) formed at the reactive K region of the molecule has been isolated from treated skin. A considerable interest has been manifested in this reactive region of the molecule and physicochemical speculations have correlated the free electron density at this site with *carcinogenic* potency

The binding of these carcinogens to protein of treated epidermis has been a subject of intensive investigation It is thought by some researchers that combinations of these carcinogens with protein may constitute an essential *first step in* carcinogenic process, proof of this has been adduced from studies of series of compounds of *varying* carcinogenic potency. The quantitative aspects of

... (IV) is a noncarcinogenic hydrocarbon, binds strongly to proteins; other investigators have found that several other noncarcinogens such as pyrene and anthracene may also bind. In any event, the nature *of the binding* and qualitative data on the nature of the protein entering into this combination is superficial. As yet this work must be considered tentative *and much* evidence would seem to discount the possibility that it provides the essential mechanism of this process.

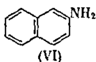


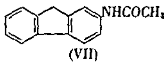
Of the azo dyes, the most representative carcinogen is *p*-dimethylaminoazobenzene (V). Similar studies have yielded information on a variety of urinary and fecal metabolites. Similarly, a series of studies has demonstrated that these dyes become bound to liver protein and that there is a quantitative relationship between the amount bound and carcinogenic potency. The studies of protein



binding in the liver preceded those in the skin with polycyclic hydrocarbons and have been carried further. However, in this instance, too, the nature of the protein involved is not yet determined and the data up to this time are still superficial.

Beta-naphthylamine (VI) induces cancer of the bladder when given by mouth to the dog, in several other species it does not have this effect. It has been found that one of the urinary metabolites of this compound, 2-amino-1-naphthol can induce cancer in the bladder of the mouse when implanted directly into the epithelium of this organ. It is, therefore, assumed that this compound is metabolized into the active carcinogen. In the instance of 2-acetylaminofluorene (VII), many tumors are induced in a variety of organs when the compound is fed by mouth; it is not active when given locally by injection The deacetylated derivative 2-aminofluorene is, however, active by injection and it is thought that the first step in the development of the direct carcinogen is the deacetylation of this compound in the gut. As with the polycyclic hydrocarbons and the azo dyes, protein-binding studies have been carried out with this compound and correlations established. Again their precise meaning awaits further studies.





In the instance of urethane (VIII), a more direct approach has been carried out than with any other carcinogen. Using a special technique of implanting the carcinogen together with a fragment of embryonic lung, it has been shown that the carcinogen does not act directly on the tissue. However, serum from a rabbit treated with urethane does contain a substance that has a direct carcinogenic action. Other evidence has been obtained that the action of this metabolite may be inhibited by concurrent dosing of the animal with orotic acid, and from this it is suggested that this carcinogen acts on pathways of formation of nucleic acid Other evidence that nucleic acids may be involved in carcinogenesis has come from the study of a series of carcinogens that have the property of cross linking; nitrogen mustard (IX) is this type of carcinogen. In this instance the mode of action of certain chemical carcinogens and of ionizing radiations are closely correlated.



As was stated at the outset, carcinogens comprise many different agents; eventually it may transpire that the various carcinogens act by many different routes and that the reaction of the tissue provoked by them has a central mechanism that may have been missed by the detailed studies of the individual agents. The difficulties in these studies arise to a large extent from the long latent period encountered between the application of the agent and the detection of the biological response. The many carcinogens now known may well not be direct mediators of the activity of special interest, in some few instances a good start has been made by the discovery of metabolites that are more potent or have a wider range of activity than the starting material. [r.s.]

Bibliography: E. V. Cowdry, *Cancer Cells*, 1955, H. S N Greene, The significance of the heterologous transplantability of human cancer, *Cancer*, 5:24–44, 1952; J. L. Hartwell, *Survey of Compounds Which Have Been Tested For Carcinogenic Activity*, U.S. Public Health Serv. Publ. 149, 1951; P. Shubik and J. L. Hartwell, *Survey of Compounds Which Have Been Tested For Carcinogenic Activity*, U.S Public Health Serv. Publ. 149, 1957, B. Sokoloff, *Cancer*, 1952, G. Wolf, *Chemical Induction of Cancer*, 1952.

Onion

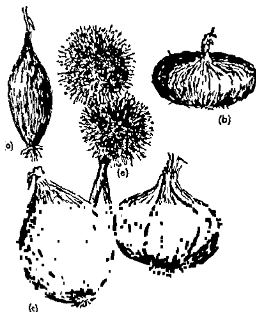
A cool-season biennial, *Allium cepa*, of Asiatic origin and belonging to the plant order Liliales. The onion is grown for its edible bulbs.

Related species are leek (*A. porrum*), garlic (*A. sativum*), Welch onion (*A. fistulosum*), shallot (*A. ascalonicum*), and chive (*A. schoenoprasum*)

Propagation. The common onion is grown as an annual and is propagated most frequently by seed sown directly in the field Onions may also be grown from transplants started in greenhouses or outdoor seedbeds, or from small bulbs, called sets, grown the previous year. Field spacing varies, plants are generally grown 1–4 in. apart in 14–18 in. rows. The Egyptian tree or top onion (*A. cepa* var. *viviparum*) produces little bulbs or topsets in the flower cluster, and the multiplier or potato onion (*A. cepa* var. *aggregatum*) multiplies by branching at the base.

Varieties. Onion varieties are classified mainly according to pungency (mild or pungent) and use (dry bulbs or green bunching) Bulb colors may be white, red, or yellow. Varieties differ markedly in their keeping quality and in their response to length of day (*see* PHOTOPERIODISM IN PLANTS) Popular dry-bulb varieties are Brigham Yellow Globe, Australian Brown, Bermuda, and Sweet Spanish Popular bunching varieties are Beltsville Bunching and White Portugal. Hybrid varieties produced from male-sterile breeding lines, are becoming more popular.

Harvesting. The harvesting of dry-bulb varieties usually starts after the leav[illegible] turn yellow



Onions (a) Oblong (b) Flat (c) Globe (d) Oblate (e) Flower heads or clusters. (*From L H Bailey, ed., The Standard Cyclopedia of Horticulture, vol. 2, Macmillan, 1937*)

and fall over generally 3–4 months after planting Bulbs to be stored are cured b...
dr
wl

...and California are important producing states The total annual farm value in the United States is approximately $55,000,000 See LILIALES, VEGETABLE GROWING [H J C]

Onion diseases. The most serious onion diseases are ...

...that grow onions commercially have diseases of some importance The weather often determines the kind and amount of disease present in any given year Wet weather at harvest time, which prevents the proper curing of the bulbs, favors the development of serious diseases, especially during transit, storage, and marketing.

Neck rot, incited by *Botrytis allii* and other species usually causes greater loss than any other disease It is responsible for considerable decay in storage onions that have not been properly cured Sometimes as much as 50% of the crop is lost.

Bacterial soft rot is probably the second most serious disease of onions. Although it occurs in the field, it causes most damage after harvest. This disease is caused by *Erwinia carotovora* and *Pseudomonas alliicola*, bacteria that are common in the soil and in water The organisms invade the moist neck of the onion at harvest time and also enter through wounds particularly under warm, humid conditions

Where cool, moist weather prevails, downy mildew caused by the fungus *Peronospora destructor* is a very destructive disease. The seedlings and leaves of the growing plants are affected.

Black mold, caused by *Aspergillus niger*, is common on onion bulbs grown in the South It seldom causes decay. The disease is readily identified by the black, powdery spores on the scales at the neck and between the outer scales of the bulb

Smudge is serious only on white varieties of onions. The causal fungus, *Colletotrichum circinans*, invades the outer scales of the bulbs and causes unsightly dark blotches. However, little damage results to the fleshy part of the bulb

Smut, a fungus disease caused by *Urocystis cepulae*, occurs in northern-grown onions. It affects the seedlings and young green onions, causing black blisters on the leaves and young bulbs.

A bulb rot caused by *Fusarium oxysporum*, a soil-borne organism, frequently causes serious damage to onions late in the season. Infections not apparent at harvest time continue to develop and cause extensive decay during storage and marketing

Other diseases, usually of minor importance, are caused by viruses, nematodes, and physiological disorders *See* NEMATODA; PLANT DISEASE; PLANT VIRUS [C.B R.]

Oniscoidea

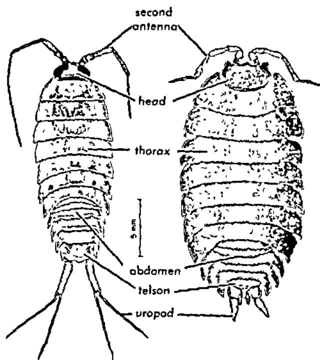
A suborder of the Isopoda which contains the terrestrial members of these crustaceans They are popularly known as sow bugs, slaters, wood lice, or, in the case of those that can roll themselves into a ball, pill bugs Land isopods commonly occur under rocks, loose bark, leaf mold, and similar moist places. They abound in humid tropical and warm-temperate regions, particularly in the Old World A few, like *Porcellio scaber* and *Armadillidium vulgare*, are practically cosmopolitan and have probably been accidentally transported by man with plants, soil, or building materials When abundant in gardens or greenhouses, they sometimes do considerable damage by gnawing on plants.

Besides being terrestrial, the Oniscoidea differ from the other seven isopod suborders in several structural characters, such as the minute size of the first antennae, the palpless mandibles, and the terminal attachment and usually styloid shape of the tail appendages or uropods.

Morphology. The body is either flattened dorsoventrally, or in pill bugs is highly vaulted like an armadillo. Its three subdivisions the head, thorax, and abdomen, are broadly joined The lateral ...

... surface may be smooth or variously sculptured with tubercles, ridges, or spines. Setae vary in kind and quantity. Adults range from 0.5 to 3.0 cm in length. Sexual dimorphism is rare

Six fused segments, including the first thoracic somite, comprise the head. It bears two pairs of antennae (the first being vestigial), two sessile



Representative oniscoideans

compound eyes, and four pairs of mouthparts, the mandibles, first and second maxillae, and maxillipeds

The thorax has seven free segments, each bearing a pair of similar seven-jointed walking legs. In females, the bases of the first five pairs of legs give rise to thin lamellae which overlap to form a brood pouch beneath the thorax

The abdomen comprises five segments plus a terminal telson. Its appendages are biramous and include five pairs of platelike pleopods and one pair of uropods. In males, the inner branches of the first two pairs of pleopods are transformed into copulatory stylets.

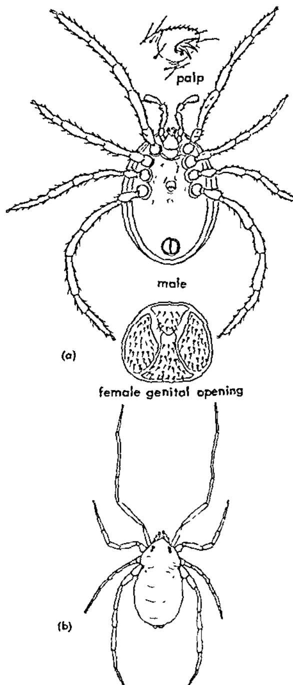
Phylogeny. The Oniscoidea doubtless stem from ancient marine isopods and have evolved to land life along several lines Their success in invading terrestrial habitats depends largely on adaptations for both aerial respiration and water regulation. Specializations of the respiratory pleopods provide the clearest relationship to their ecological distribution, but cutaneous glands, the integument, excretory organs, and physiological mechanisms are also variously involved in coping with terrestrial problems.

The primitive Ligiidae and Trichoniscidae represent transitional stages as their essentially branchial pleopods and moisture requirements limit them to damp littoral, halophilic, or riparian habitats. More advanced families can inhabit drier environments, primarily because the outer branches of two or more pairs of pleopods contain ramifying air pockets, called pseudotracheae or "white bodies," which assist atmospheric respiration But even so-called desert species with well-developed pseudotracheal systems are imperfectly land adapted. All terrestrial isopods require ecological niches with high microhumidity or with available free water, since they cannot survive exposure to unsaturated air except for short periods See ISOPODA; TERRESTRIAL ECOSYSTEM. [M.A.M]

Bibliography: E. B. Edney, Woodlice and the land habitat, *Biol. Revs.*, 29(2):185–219, 1954. A. Vandel, Essai sur l'origine, l'évolution et la classification des Oniscoidea (isopodes terrestres), *Bull. biol. France et Belg.*, Suppl. 30, 1943; W G Van Name, *The American Land and Fresh Water Isopod Crustacea*, Am. Museum Nat. Hist. Bull 71, 1936.

Onychopalpida

The smallest and most primitive of the five suborders of the Acarina, comprising two widely divergent families, the Opilioacaridae and Holothyridae. As a group, they are characterized by the



Onychopalpida. (a) A holothyrid mite (b) An opilioacarid mite. (The Institute of y, University of Maryland)

possession of claws on the palpal tarsus that are paired, like those on the legs, or that are variously fused and modified They are also unique in that they possess four pairs of ventrolateral or dorsolateral stigmata, or breathing pores, but no associated peritremes *Peritremes are chitinous tubes* that are associated with the respiratory structures, the stigmata and trachea.

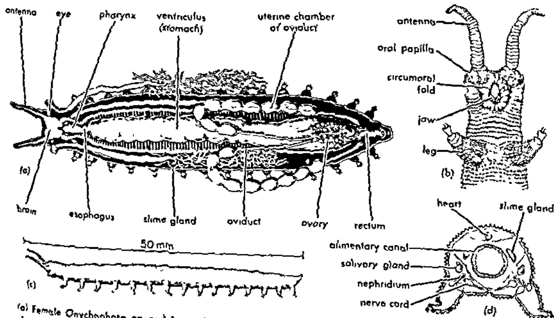
The Opilioacaridae are moderately large mites, 1-2 mm in length. They have long legs, a leathery cuticle, and indications of segmentation on the dorsal surface of the hysterosoma The hysterosoma is the posterior region of the body extending from *the region of the third pair of legs* to the terminus of the body They look like small phalangids, to which they may be closely related They live under stones and other debris and probably prey upon *other small arthropods* They occur in the Mediterranean area, the southern United States, and the West Indies.

The Holothyridae are among the largest of the mites, reaching 2-7 mm in length They are hemispherical in shape, though longer than broad, and have a deep brown, smooth, heavily sclerotized cuticle. They are probably predaceous in habit They have been found in the Indo-Australian region and on the island of Mauritius, where it has been reported that an irritant poison they secrete causes the death of ducks and geese that swallow them *See* ACARINA. [J H C]

Onychophora

A phylum of unusual wormlike animals a few inches long found in various tropical and subtropical parts of the world. These creatures are of particular interest to zoologists because they combine features of both the annelid worms and the arthropods They are neither worms nor arthropods, however, and are given a phylum of their own, the *Onychophora*. These animals have a double row of short, thick legs along the length of the body and a pair of large tentacles, or antennae, on the head. They live on the ground, always in damp places, and feed on small arthropods. The first onychophoran found was named *Peripatus*, "the walker," to distinguish it from the crawling or swimming worms Most of the other genera have similar names, such as *Eoperipatus*, *Paraperipatus*, *Peripatopsis*, *Ooperipatus*, and *Opisthopatus*.

The head of an onychophoran is not distinct from the body, and the body of the adult is not segmented. The embryonic development, however, shows that the Onychophora are fundamentally segmented animals, as are the annelid worms and the arthropods The entire trunk is closely ringed with transverse bands of small, spine-bearing tubercles. The head bears the antennae anteriorly and a small simple eye on each side. The mouth is on the undersurface in an oval depression surrounded by a lobulated circumoral fold A pair of strong, 2-hooked jaws converge from the sides beneath the mouth. The jaws are not equivalent to the mandibles of an arthropod They are the claws of a pair of otherwise reduced limbs The mandibles are formed from the basal segments of a pair of appendages On each side of the head a short appendage, known as the oral papilla, gives exit to the duct of a much-branched slime gland in the body. When the animal is irritated or disturbed, the slime may be ejected to a distance of several inches, serving as a means of defense The legs are thick, tapering, lateroventral outgrowths of the body wall, and are also ringed with tubercles Each leg ends in a small, 2-clawed, foot lobe



(a) Female Onychophora opened from above (b) Undersurface of head and front of body (c) Side view (d) Diagrammatic cross section of body showing position of internal organs.

The respiratory organs are bunches of fine tracheal tubules that open from pores scattered over the surface of the body. The tracheal system of the Onychophora is thus quite different from that of terrestrial arthropods.

The alimentary canal is a wide tube that extends straight through the body from the mouth to the anus. A pair of long, tubular, salivary glands opens into it behind the mouth. Excretion is effected by both the alimentary canal and a double row of simple nephridia opening at the inner sides of the legs.

The nervous system includes a well-developed brain which gives off nerves to the antennae and eyes, and a subesophageal ganglion, in the lower part of the head, from which extend two widely separated ventral nerve cords connected by numerous transverse commissures. The nerve cells of the cords are not condensed into ganglia as they are in the arthropods, and the body nerves are given off directly from the cords.

A pulsating dorsal blood vessel, or heart, keeps the blood in circulation The blood enters the heart through lateral apertures and is discharged from the anterior end, to flow back through the body cavity.

The reproductive organs, the testes of the male and ovaries of the female, lie in the posterior part of the body above the intestine The oviducts loop forward, and then turn backward, to come together in a common median opening near the end of the body. The oviducts of viviparous species are enlarged in a series of uterine chambers in which the embryos develop. In some species, the male inseminates the female by attaching numerous small sperm-containing capsules, called spermatophores, almost anywhere on the outside of her body. The integument beneath each spermatophore then disintegrates, allowing the spermatozoa to enter the body cavity and swim to the ovaries, to penetrate and fertilize the eggs within *See* ANIMAL KINGDOM.

[R.E.S]

Onyx

The name onyx is applied correctly to banded chalcedonic quartz, in which the bands are straight and parallel, rather than curved, as in agate. Unfortunately, in the colored-stone trade, gray chalcedony dyed in various solid colors such as black, blue, and green is called onyx, with the color used as a prefix. Because the color is permanent, the fact that it is the result of dyeing is seldom mentioned

The natural colors of true onyx are usually red or brown with white, although black is occasionally encountered as one of the colors. When the colors are red-brown with white or black, the material is known as sardonyx; this is the only kind commonly used as a gem stone. Its most familiar gem use is in cameos and intaglios, in which the figure is carved from one colored layer and the background in another. *See* CAMEO; CHALCEDONY; GEM; INTAGLIO (GEMOLOGY); QUARTZ.

[R.T.L.]

Oogenesis

The processes of egg formation by which certain cells, the oogonia, of the ovary enlarge and undergo meiosis (*see* GAMETOGENESIS). When the nucleus of a terminal oogonium begins to undergo its meiotic changes, the cell is designated a primary oocyte. The nuclear changes proceed to the tetrad stage early and remain in that condition until the oocyte is fully grown. As the oocyte enlarges, the nucleus also expands and is designated a germinal vesicle. The two meiotic divisions, however, do not occur until full growth is attained. In most species of animals these divisions are not completed until after both ovulation, which is the release of the oocyte from the ovary, and fertilization In two groups of animals, coelenterates and echinoids, meiosis is generally completed intra-ovarially. In most mammals sperm entry occurs just after the first meiotic division, known as the secondary oocyte stage

The first meiotic division of the oocyte results in a large cell, the secondary oocyte, and a small structure called the first polar body. The second meiotic division again produces a large cell, the ootid, and a small second polar body At the same time, the first polar body divides, although sometimes it fails to do so. The result is four cells, each with the haploid number of chromosomes, but only the ootid is functional, as the egg, while the polar bodies degenerate. In species in which the spermatozoon enters a primary or a secondary oocyte, it waits for completion of the meiotic divisions before fusing with the egg nucleus *See* EMBRYOLOGY, EXPERIMENTAL.

Submerged meiosis. It is possible, by experimental means such as heat-treatment, to induce the oocyte to undergo submerged meiotic divisions so that two, three, or four haploid nuclei remain in the ootid, and two, one, or no polar bodies are formed These egg nuclei can fuse with the haploid sperm nucleus and thus give rise to triploid, tetraploid, and pentaploid individuals Such polyploid animals have been produced and studied extensively in amphibians. In general, their cells are of correspondingly large size, but since the animal is of normal size there are correspondingly fewer cells The learning ability of a triploid salamander is found to be inferior to that of the normal diploid. This would seem to be related to its possessing about two-thirds of the normal number of neurons (*see* CELL CONSTANCY)

Production of oogonia occurs by mitotic cell divisions throughout life, in most species of animals However, in mammals it generally ceases early. In humans this occurs shortly after birth, when there are some 100,000 oogonia present in the ovary They do not proceed beyond an early oocyte stage until puberty. From then, for approximately 30 years, a single fully formed egg, but occasionally 2–5, is produced each month Many more may start to grow and then degenerate. This process of regression is termed atresia, and tic follicles con-

taining degenerating oocytes are commonly seen in the human ovary at all times.

Growth of the oocyte takes place, as noted above, before it has undergone the first meiotic division and, therefore, while it is still under the influence of the diploid set of chromosomes. This influence is sometimes manifest in the phenomenon of maternal inheritance (*see* GENETICS) An example is the inheritance of the direction of coiling of the shell in certain fresh-water snails (*Limnaea peregra*), which is determined by the genes present in the oocyte rather than by those provided to the zygote by union of the haploid egg and sperm pronuclei upon fertilization

Eggs of all animals undergo a considerable enlargement during the transformation of the terminal oogonium into the fully formed oocyte The increase in mass ranges from some thousandfold for animals such as mammals, with small eggs, to many billions of times for animals such as birds and sharks, that have large eggs The increase in mass represents accumulation of reserve food materials, largely in the form of lipids and proteins, that are later utilized during development This food reserve is termed yolk and is mainly in the form of small, spheroidal bodies It is often thought that the deposition of yolk signifies considerable synthetic activity on the part of the growing oocyte However, experiments designed to explore this, by use of radioactive tracers and other means, now show that most of the accumulated materials are fully formed, or nearly so, when supplied to the oocyte, their synthesis having taken place in other tissues of the body

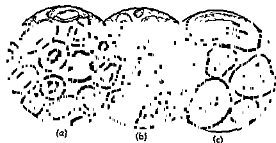
Gonadotrophic hormones. The hormones, produced by the anterior lobe of the pituitary gland in vertebrates, control the ripening of the oocyte and ovulation Removal of the anterior pituitary results in atrophy of the ovary Atrophy can be prevented by injection ...
ture anim...
can induc-
oocytes, and their ovulation In mammals, there is good evidence for the existence of at least two distinct pituitary hormones that affect the ovaries. One, termed the follicle-stimulating hormone (FSH), causes ripening of the egg and its surrounding Graafian follicle The other, the luteinizing hormone (LH), induces ovulation and growth of the cells lining the empty follicle into a structure known as the corpus luteum

Periodicity of oogenesis. This is exhibited by most animals, as represented by the occurrence of definite breeding seasons In many animals, this synchronization of reproductive activity is achieved, as in plants, by a physiological effect of increasing, or of decreasing, day length Experiments among vertebrates have shown that the triggering action of changing day length is mediated through the pituitary gland Many invertebrates, like certain annelids and mollusks, and some vertebrates, such as the grunion on the Pacific Coast, ripen and release their gametes at particular phases of the moon Others have internal "clocks" that control the time of oogenesis In primates, including humans, the typical 28-day cycle of ovulation is independent of lunar phase but may well have originated in ancestral mammals whose reproductive activity was synchronized with particular phases of the moon. *See* PERIODICITY IN ORGANISMS. [A.TY.]

Oolite and pisolite

Oolites are small, more or less spherical particles commonly found in limestones and dolomites. Most oolites are 0.5–1.0 mm in diameter, but their size range is much greater. Pisolites are similar particles that are greater than 2.0 mm in diameter. Oolites show varying degrees of departure from sphericity; some may be ellipsoidal, others may be appreciably flattened or distorted. The term oolite has been used to denote both the small, spherical bodies and the rock composed of an aggregate of these bodies Some geologists prefer to call the particles oolite and the rock by its common lithologic name, prefixed by the word oolitic, for example, oolitic limestone.

Sectioned oolites show ...
structures or b...
are of materia...
frequently the ...



Oolitic limestones. (a) *Pleistocene oolites, Great Salt Lake, Utah* Diameter 3 mm Oolites *consist of subangular detrital quartz grains enclosed by carbonate having both concentric and radial* fibrous structure. Radial *fibrous carbonate* is calcite, at least some of the concentric carbonate *(right center and top)* is aragonite *An incipient cement* composed of finely granular *calcite rims the oolites, but rock is very po*rous (b) Oolitic limestone, Volksen, Deister Mountains, Germany *Diameter 3 mm.* Oolites consisting of shell fragments encased by *microcrystalline calcite* (dark stippling) are firmly cemented *by a matrix* of finegrained calcite having *somewhat variable* grain size. (c) Composite *oolites* (Pleistocene), Pyramid Lake, Nevada Diameter 6 mm Large calcareous *oolites* consisting of cryptocrystalline (stippled) and *radial* fibrous (clear) concentric *layers* Fibrous layers are calcite, cryptocrystalline layers are at least partly aragonite. Nuclei are *fragments* of broken oolites, clusters of tiny oolites (right and *center*), and bits of granular carbonate (lower right). Incipient cementation as in *a* (From *H. Williams, F. J. Turner, C. M. Gilbert, Petrography, An Introduction to the Study of Rocks in Thin Sections,* Freeman, 1954)

quartz grains. The appearance of oolites suggests that they have grown outward from the core by successive precipitations of calcium carbonate in thin concentric shells. Although oolites may be composed of many materials, mainly calcite, aragonite, silica, hematite, and dolomite, by far the most common in the geologic column are the calcareous ones. Siliceous and dolomitic oolites are formed by the replacement of an original calcareous oolite. Phosphatic and hematitic oolites seem to have formed as primary oolites. The explanation for the origin of oolites generally given is that they represent inorganic precipitation in turbulent waters, where the small grains roll with the current as they gradually pick up more and more layers of precipitate. *See* CALCARENITE; CHERT; DOLOMITE; LIMESTONE; SEDIMENTARY ROCKS. [R.SI.]

Opal

A natural hydrated form of silica. There are many different varieties of opal, but the best known are those which are highly prized as gem stones. Precious opal displays the property of opalescence, a fine play of spectral colors resulting from the interference of light rays within the stone. Fire opal shows intense orange-to-red reflections against a yellow-to-orange body color Black opal has a black background against which the colors are displayed. Common opal is milk-white, yellow, green, or red, but without opalescence. The variety hyalite is clear and colorless with a globular surface Fine precious opals are found in Hungary, Mexico, Honduras, and New South Wales, Australia In the United States opal has been found in Nevada and Idaho.

Opal is amorphous and usually occurs in botryoidal or stalactitic masses It has a conchoidal fracture and hardness of 5–6 (Mohs scale) The specific gravity varies from 1.9 to 2.2, depending upon the water content Opal is found in cavities in igneous and sedimentary rocks and in fossil wood in which it is the petrifying substance. As geyserite or siliceous sinter it is deposited from geysers in Yellowstone National Park. Its largest deposits are in sedimentary beds as diatomite which result from the accumulation of tiny opalean tests of diatoms. Such a deposit at Lompoc, California, is 400 ft thick *See* DIATOMACEOUS EARTH, GEM; SILICATE MINERALS; SILICEOUS SINTER. [C.S.HU.]

Opaque medium

One which is impervious to rays of light, that is, not transparent to the human eye. By extension, a medium may be described as opaque if it does not transmit infrared waves or other regions of the electromagnetic spectrum, such as the x-ray, ultraviolet, and microwave regions. The property of zero transmittance does not necessarily imply total reflectance; that is, opacity can result both from reflection and from absorption of incident rays *See* ABSORPTION (ELECTROMAGNETIC RADIATION). [M.C.M.]

Open circuit

A condition in an electric circuit in which there is no path for current between two points. Examples of open circuits are a broken wire and a switch in the OPEN, or OFF, position. *See* CIRCUIT, ELECTRIC.

Open-circuit voltage is the potential difference between two points in a circuit when a branch (current path) between the points is open circuited. Open-circuit voltage is measured by a voltmeter which has a very high resistance (theoretically infinite), such as a vacuum-tube voltmeter. [C.F.C.]

Open systems, thermodynamics of (biology)

Thermodynamics is founded on the basis of two postulates, the first and second laws. From these laws are derived the relations that may obtain among those macroscopic parameters necessary to describe a system. The parameters include energy, mass quantities, volume, pressure, temperature, electrical potential, and electrical charge. Thermodynamics is applied to discrete portions of the universe, termed the system, and bounded from the remainder of the universe, termed the environment, by definitive limits. The definitive boundaries may have one or more of a number of special properties; for example, an isolated system may exchange neither energy nor mass with its environment, and an adiabatic system may not exchange heat with its environment An isolated system will ultimately approach a state of equilibrium. Thermodynamics deals only with such states of equilibrium For thermodynamic purposes, changes in state of the system are conceived to be carried out reversibly in successive states of equilibrium Any extension of thermodynamics to nonequilibrium processes must therefore invoke nonthermodynamic assumptions *See* THERMODYNAMIC PRINCIPLES; THERMODYNAMIC PROCESSES.

Biological systems. Biological systems are open systems in the thermodynamic sense Across the boundaries of open systems both matter and energy may pass. Thus the biological system may take in certain matter as foodstuff which on undergoing a series of chemical transformations is converted to matter and eliminated as waste products From such chemical reactions the biological system derives its energy, thus permitting it to perform its internal functions of growth and maintenance and to perform work on the environment For a rigorous application of thermodynamics to biological systems it is necessary that biological processes be carried out reversibly in successive states of equilibrium This applies to the system in relation to its environment as well as to all the processes within, namely, the scalar chemical reactions and vectorial transport reactions *See* BIOLOGICAL OXIDATION; CELL (BIOLOGICAL)

Biological systems, however, are characterized by the fact that they are not equilibrium systems

It is convenient, therefore, in treating biological systems to apply that extension of classical thermodynamics known as the thermodynamics of irreversible processes. As a theoretical basis for this extension, L. Onsager drew a parallel between the rate of regression of a statistical fluctuation of a variable about equilibrium and a macroscopic irreversible change in this variable. This established general validity for the theory when restricted to systems whose departure from equilibrium is small. Inherent in this extension is the implicit assumption that the laws of thermodynamics are valid outside of equilibrium.

Theory. The more cogent points, together with the necessary assumptions involved in the thermodynamics of irreversible processes, are here presented.

There exists a function, S, termed the entropy of a system, which is a function of the state of the system. This function is extensive in that the total entropy is a sum of the entropies of the parts of the system. The entropy of a system may change by interaction of the system with its environment d_eS and by internal changes in state of the system d_iS. Thus

$$dS = d_eS + d_iS \tag{1}$$

The entropy change d_eS is related to the heat q absorbed from the environment by

$$d_eS = \frac{q}{T} \tag{2}$$

where T is the absolute temperature. Furthermore,

$$d_iS > 0 \tag{3}$$

for all natural changes, and

$$d_iS = 0 \tag{4}$$

for all reversible changes. It follows, therefore, that

$$dS > \frac{q}{T} \tag{5}$$

for all natural processes. Equations (1) through (5) comprise a statement of the second law of thermodynamics. See ENTROPY.

It is assumed that the thermodynamic state of a system, and consequently the entropy, may be completely described at any instant in time by a set of extensive macroscopic variables. Such variables include, for example, the mass quantities of each chemical constituent, the volume of the system, the electrical charge, and the energy, and are here designated as A_i, with equilibrium values, A_i^0. This implicitly assumes that the entropy is not explicitly a function of time, and that the thermodynamic relations are applicable to systems departing from equilibrium. It is further assumed that the entropy of the system may be expressed in terms of these variables by a Taylor's series expansion about equilibrium.

$$S - S^0 = \Delta S = -\sum_i \frac{\partial S}{\partial \alpha_i}\alpha_i - \frac{1}{2}\sum_{ij}\frac{\partial^2 S}{\partial \alpha_i\,\partial \alpha_j}\alpha_i\alpha_j + \cdots \tag{6}$$

where $\alpha_i = A_i - A_i^0$. Since the entropy is a maximum at equilibrium, the first term on the right-hand side vanishes. If only quadratic terms are retained, the rate of entropy production in the system due to irreversible processes is

$$\frac{d\,\Delta S}{dt} = \dot{S}(\alpha_i\alpha_j) = -\sum_{ij}\frac{\partial^2 S}{\partial \alpha_i\,\partial \alpha_j}\alpha_i\frac{d\alpha_j}{dt} = -\sum_{ij} g_{ij}\alpha_j\dot{\alpha}_i \tag{7}$$

The rate of change with time of the variables α_i is taken as the fluxes $J_i = \dot{\alpha}_i$. The thermodynamic forces are then

$$X_i = -\sum_{ij} g_{ij}\alpha_j$$

Consequently, the rate of entropy production becomes

$$\dot{S}(J_iX_i) = \sum_i J_iX_i \geq 0 \tag{8}$$

and is seen to be a sum of products of forces and fluxes. The inequality is valid for all natural processes occurring in the system, whereas the equality is valid for all reversible changes.

For processes occurring near equilibrium, linear phenomenological equations may be assumed. Thus the set of equations

$$J_i = \sum_j L_{ij}X_j \quad (i = 1, 2, \ldots) \tag{9}$$

may be written in which L_{ij}, $i = j$, is the coefficient relating the flux J_i to its conjugate force X_i, whereas the coefficient L_{ij}, $i \neq j$, describes a coupling of the flux J_i to the force X_j. Onsager has given statistical validity to reciprocal relations among the Ls, $L_{ij} = L_{ji}$, valid for processes without significant inertial or magnetic forces. This type of force is excluded for systems herein discussed.

Equations (8) and (9) may be combined to give

$$\dot{S}(X_iX_j) = \sum_{ij} L_{ij}X_iX_j \geq 0 \tag{10}$$

a form, quadratic in the forces, and positive definite. Thus the coefficients L_{ij} are the elements of a symmetrical positive matrix.

A dissipation function ϕ may be defined as

$$\phi(X_iX_j) = \tfrac{1}{2}\sum_{ij} L_{ij}X_iX_j \tag{11}$$

and is seen to be one-half the rate of entropy production given by Eq. (10). The expression $S(J_iX_i)$ given in Eq. (8) is linear in the Xs and positive definite. The dissipation function is quadratic in the Xs and positive definite. The difference, $S(J_iX_i) - \phi(X_iX_j)$, is therefore a maximum if

one adopts the convention that only the Xs are to be varied. Thus

$$\sum_i J_i X_i - \tfrac{1}{2} \sum_{ij} L_{ij} X_i X_j = \max \qquad (12)$$

and a variation gives

$$\delta\left[\sum_i J_i X_i - \tfrac{1}{2} \sum_{ij} L_{ij} X_i X_j\right] = 0 \qquad (13)$$

$$\sum_i [J_i - \sum_j L_{ij} X_j]\, \delta X_i = 0 \qquad (14)$$

Since the δX_is are independent variations, the extremum condition demands that their coefficients vanish, thus giving the set of linear relations of Eq. (9). Embodied in this variation principle of Onsager are also the reciprocal relations $L_{ij} = L_{ji}$. Onsager designates this variation principle as an extension of Rayleigh's principle of least dissipation. A system in undergoing change from one thermodynamic state to another does so along that path which involves the least dissipation of energy It may be shown that this path in time is that described by a set of first-order differential equations, which as a consequence of Eq. (14) represents a maximum rate of decrease in the rate of entropy production.

From the definitions of the forces X_i as first utilized in Eq. (8), the inverse relation

$$\alpha_i = -\sum_j C_{ij} X_j \qquad (15)$$

may be obtained wherein the elements C_{ij} are of a matrix inverse to that of $[g]$. Taking the time derivative of Eq (15) gives

$$J_i = \dot{\alpha}_i = -\sum_j C_{ij} \dot{X}_j \qquad (16)$$

which in conjunction with Eq. (14) gives a deterministic set of first-order linear homogeneous differential equations in the Xs. Thus

$$\sum_j [C_{ij} \dot{X}_j + L_{ij} X_j] = 0 \quad (i = 1, 2, \ldots) \qquad (17)$$

The general solution is given by

$$X_i = \sum_j A_{ij} C^{-\beta_j t} \quad (i = 1, 2, \ldots) \qquad (18)$$

where the β_js are the characteristic roots of Eq. (17) and the A_{ij}s are constants of integration determined by the boundary conditions

In a system of n forces it is seen that, without any imposed external constraints, the approach to a state of equilibrium involves a time course composed of n exponential processes for each of the Xs. Similarly if, in a system of n forces, k constraints are imposed, that is, if k of the forces are held constant by having the system in contact with an infinite reservoir for such quantities as heat or a number of the chemical constituents, the approach to the steady state for each unrestrained X will involve a time course composed of $n - k$ exponential processes. The resulting steady state is said to be of the kth order and represents a state with a minimum rate of entropy production in which all fluxes whose conjugate forces are not subject to external constraint must vanish. With this designation true equilibrium is that of a zeroth-order steady state.

This completes the development of the more general and significant aspects of the thermodynamics of irreversible processes, valid for all systems whose departure from equilibrium is not large, and for systems in which forces of inertial or magnetic origin are specifically excluded. Inclusion of such forces would require modification of the Onsager reciprocal relations. Fundamental to this theory is a variational principle which represents a principle of least dissipation.

Application. Biological systems are extremely complex as living units, and it is generally necessary in their study to introduce certain simplifications. For some purposes, it may be assumed or actually specified that the system be at constant temperature and pressure, whereupon it may be shown that the rate of entropy production is equal to the rate of decrease in the Gibbs free energy for the system. For other purposes it may be convenient to assume that the system is homogeneous in its extent, and thereby all processes may be conceived and written in terms of simple chemical reactions. Still, for other purposes, the boundaries of the system may be so specified as to reduce those considerations involving the exchange of matter between system and surroundings to only a few chemical constituents It is only with simplifications of this kind that thermodynamics of irreversible processes as herein developed can be reasonably applied Many fields of application and investigation present themselves. A central problem is that of the transduction of energy in biological systems, whereby via sequential reactions of a highly organized nature chemical energy is transformed for use in all cellular processes.

In rather qualitative terms some of the above concepts may be illustrated by a living system in contact with its environment It may be specified that heat and energy as well as certain matter may pass across its boundary. Thus the cell is allowed to gain such chemical entities as glucose and oxygen and to eliminate others such as carbon dioxide and water. The complete oxidative metabolic conversion of glucose to carbon dioxide and water involves a large number of chemical reactions. The reactions are sequential in extent with many subsidiary branchings Sequences may be cyclic with consequent regeneration of the intermediate constituents involved They are subject to multiple regulatory restraints arising from internal mechanisms and also from environmental influences such as hormones. Included in such a complex set of reactions are certain intermediates. the high-energy phosphates. These substances appear to be more directly utilized as energy sources for the many cellular processes, among which must be included the maintenance of cellular integrity and homeostasis. This maintenance of the internal environment of the cell which ensures proper conditions for the many essential metabolic reaction [illegible] s specialized

ansport mechanisms at the cellular boundary hese transport systems, operating in a selective ıanner, serve to bring about a controlled exchange f matter with the environment. It may be assumed hat the environment is subject to experimental manipulation and control. If, then, at constant temperature and pressure the potential of all those constituents with which the cellular system may exchange is held constant in the environment, a steady state is ultimately approached In this state the constant rate of entropy production arising from the irreversible processes occurring is at a minimum The potentials of all constituents within the system remain constant in time Many cells, tissues, and organisms in their relatively dormant or resting states approximate such steady behavior Thus a muscle cell at rest exhibits relatively stationary behavior, oxidizing glucose at a constant rate All of the energy derived from this oxidation is dissipated at a constant rate The transient or nonsteady states are in many ways more interesting These may be conceived of as variations about some given *steady* state. Thus certain stimuli, representing a definite change in environmental conditions, may lead to *transient* behavior In a muscle cell, for instance, an appropriate single electrical stimulus causes an excitation with a resultant contractile twitch This is followed by a gradual return to the original steady conditions During such a transient state an increased rate of entropy production occurs, and with proper mechanical linkage, thermodynamic work may be performed by the system

Value in the application of thermodynamic principles to biological systems lies in the fact that such theoretical considerations serve as sound and fundamental guides in concept, approach, and conclusion in the search for greater understanding of living systems *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY, MUSCLE (BIOPHYSICS) [F M S]

Bibliography S R de Groot. *Thermodynamics of Irreversible Processes*, 1951; K G Denbigh *Thermodyna...*

... Reciprocal relations in irreversible processes, *Phys Revs*, 37(4) 405–426, 1931; I Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, 1955

Open-loop control system

A control system in which the system outputs are controlled by the system inputs only In such systems no account is taken of the actual system output This is in contrast to closed loop control systems, in which the system output is controlled by some combination of the system input and output For a discussion of all references to closed loop control systems, see CONTROL SYSTEMS

The block diagram of Fig 1 shows the general configuration of an open loop control system. A simple example of an open-loop control system is the electric-light control system in the home. In this system, the presence or absence of light from a lamp is controlled by the position of an electrical switch Using the terminology of Fig. 1, the position of the switch (whether it is on or off) is the controlling input, the light bulb in the lamp is the controlled system, and the light produced by the light bulb is the controlled output.

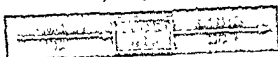


Fig. 1. Open-loop control system.

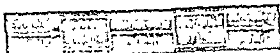


Fig 2. Open-loop control system with power amplification.

Classification ...

open-loop control system without power amplification is one in which all the output power is supplied by the controlling input. The ordinary household thermometer is an example of an open-loop control system without power amplification. The controlling input is the temperature of the surroundings, the controlled system is the mercury column, and the controlled output is the height of the mercury column. An increase of temperature causes the mercury column to increase its length. The power required to lengthen the mercury column is directly obtained from the temperature of the ...

... position of an on-off electrical switch that energizes the motor, the controlled system is the grinding wheel, and the controlled output is the velocity of the surface of the grinding wheel. The power required to move the grinding wheel does not come directly from the controlling input but from some intermediate device (the electric motor), which is directly controlled by the switch. Although the block diagram ...

... to indicate the existence of power amplification If the terminology of Fig 2 is applied to the grind...

... wheel), and the controlled output is the surface velocity of the grinding wheel. *For additional examples of open-loop control, see* ALARM SYSTEMS; CLOCK CONTROL SYSTEMS; REMOTE CONTROL SYSTEM; REPEATER, SYNCHRO; TELEMETERING; TRAFFIC CONTROL SYSTEMS

Advantages. Open-loop control systems are, in general, considerably simpler than closed-loop systems. This simplicity usually results in a more economical control system. A more important advantage of open-loop control is the elimination of stability problems, which exist in many closed-loop systems.

Disadvantages. In open-loop systems, the controlled output is determined only by the controlling input and the dynamic characteristics of the controlled system. Therefore, the dynamic characteristics of the system must be accurately known if effective control is to be achieved. In an electrically-heated oven the amount of current passing through the heating coils is adjusted by means of a potentiometer. As the current is increased, the power supplied to the oven and the internal oven temperature are increased. It is possible to measure the oven temperature resulting from each dial setting of the potentiometer. Once this is done, it is no longer necessary to measure the oven temperature as long *as all conditions remain unchanged. It* is only necessary to set the potentiometer dial to the setting corresponding to the desired temperature. If, however, the thermal properties of the oven are changed by leaving the oven door open, so that more heat is lost than when the potentiometer dial was calibrated, then a particular potentiometer setting will no longer correspond to the same oven temperature and a new oven calibration must be made under these new conditions. Such continual recalibration is tedious and expensive

Closed-loop control systems exhibit the property of self-calibration; that is, there is automatic compensation for changes in system characteristics. It is also possible to minimize the effects of external disturbances in a closed-loop system, whereas in an open-loop system no account is taken of such disturbing inputs. An example of an external disturbance in the electric oven is a variable breeze passing over the surface of the oven As the velocity of the breeze varies, more or less heat is carried away. This has an effect on the oven temperature, but is not taken account of at the input (the dial setting). It is primarily for the two reasons above that the majority of control systems in existence are closed-loop in nature. *See* CONTROL SYSTEMS

[J.C.TR.]

Bibliography: *J. C. Gille, M. J. Pélegrin, and* P. Decaulne, *Feedback Control Systems*, 1959.

Opera glasses

Small binocular telescopes adapted for use where magnification and field of view are secondary to compactness and cost. The Galilean design is most often used in opera glasses because of its inherent shortness and simplicity of construction (*see* TELESCOPE). The principal deficiency of this design is the location of its exit pupil forward of the eyepiece, which restricts the field of view to less than one-half of the field of a prism-erecting binocular of equal power. Because of this restriction, opera glasses

Operations research

An organized and systematized study of complex situations such as arise in the activities of risk-taking organizations of people and resources. Business decisions and military activities are important examples of situations studied by operations research. Such study uses a specific disciplinary approach

Objectives. The purpose of operations research is to provide, on a continuing and regenerative basis, more complete and explicit understanding of complex situations and thus to supply knowledge for more rational and systematic objective-setting and decision-making and to lead toward more effective joint performance of individuals in such organizations. Such understanding may include description of six factors

Assumptions. Underlying one's understanding of a situation are assumptions in respect to both its external environment and its internal composition *Other assumptions pertain to the objectives* and purposes to be obtained in respect to the situation, the description of the elements in the situation or impinging upon it, and the interrelation of these elements.

Results The actions, the relationships required, and the degree of commitment needed to achieve theoretically the desired results need description together with the theoretical limits on such results.

Dynamic range. Descriptions are required of the theoretical range of responses of the situation to conceivable volitional actions of people, activities outside the situation but related to it, or interactions within the situation.

Critical factors. In a situation, changes of some factors in magnitude, intensity, frequency, or configuration, dictate a need for change in the situation itself (either in objectives and purposes or to maintain existing objectives and purposes). Measures are required for such factors to observe significant changes in them as compared with expectations

Classification of situations. Operations research classifies situations as to their importance to the achievement of the over-all objectives and purposes of the total organization.

Interrelations The relation of individual situa*tions to each other and to the objectives* and purposes of the whole organization needs description together with the impact of such situations on the performance of the whole.

The type of situation in which operations research appears to be most useful is that where (1) sufficient elements, or elements of such character, are found so that their relationship is not intuitively or easily discerned. (2) all of the elements may not be known, or the elements may behave in a not readily predictable manner, or, in general, knowledge is incomplete and judgment is involved, (3) there can be alternatives for the allocation of effort and related and resultant effective-

-known includ-
y peo-
ere is
need for joint performance among several or many individuals to obtain desired objectives that requires a degree of common vision or mutual understanding of the situation to achieve the desired performance

Operations research frequently studies situations involving only some of these conditions. An example is a situation containing many elements, but about which it is assumed that the properties of the elements are given or are predictable within stated and fixed limits. The purpose under these circumstances may be to obtain an optimal allocation of effort and related and resultant effectiveness. In these cases it is either explicitly or tacitly assumed that the conditions exist in isolation or are maintained as a matter of policy from outside the situation so that the answers found are applicable only within the stated and fixed limits

Subjects. The subject matter of operations research includes, broadly, the activities of risk-taking organizations of people and resources. Operations research makes certain assumptions about these organizations and their activities. It assumes that the elements of such activities may be studied as if they were systems. Numerous and significant properties may be described concisely because, in a broad sense, they are capable of being arranged in an orderly fashion. The order may be repetitive, structural, or on any basis that is capable of formal description. Activities within such organizations are basically rational in that they contain choices, are subject to logical treatment, and profit from the application of information and hypothesis derived from theoretical and empirical study. Such organizations to which this discipline applies are not physical, but socioeconomic. Their purpose is the commitment of present resources to future expectations. They exist principally to make and take risks and thereby to change the environment in which they exist and in turn to adapt themselves to changes in the environment

In dealing with these organizations, operations research assumes further that all observable phenomena, their periodicity, regularity, and relations are knowledge at best about past or present conditions and are subject to change by human action within or without the organization. Statements about this behavior other than purely historical statements, are assumed to be statements of expectations. Therefore valid statements must include (1) statements of the assumptions underlying expectations, (2) measurements needed to test the validity of these assumptions continuously or to detect changes that might invalidate them, and (3) statements of what might be done to change the regularity, periodicity, and relationship of phenomena

In this reactive situation, operations research has both a narrower and a larger scope compared to physics and biology. On the one hand, all that the results can be are statements that give expectations an operational meaning. They are not statements of natural laws.

On the other hand, knowledge, within this discipline, includes identification of theoretical means to change the behavior and properties of the universe of an organization, rather than merely means to exploit and manipulate observed regularities and relationships.

Methodology. The basic methodology is that of modern scientific logic. It begins with the twofold hypothesizing or model-building stage. The first step is to make a general probabilistic statement about input and output phenomena defined in terms relevant to the situation. The second step is to make a general probabilistic statement about interrelations and interactions within the systems in respect to the input-output relationship

The second stage is the threefold one of validation exploration, and testing. In it, the first step is logical, analytical, or numerical exploration and testing of the hypothesis. Then comes experimental testing of the hypothesis as the action plan for achieving results. The final step is the regenerative one, which is the feedback from the validation and testing to the hypothesis for the purpose of refining, modifying, or changing it

There are many techniques; some are general purpose (see ALGEBRA, BOOLEAN ALGEBRA; GAME THEORY, LOGIC, MATRIX THEORY; PROBABILITY; SET THEORY, STATISTICS). Others are special purpose (see LINEAR PROGRAMMING; QUEUING THEORY) applicable within this discipline.

Nature of results. The nature of expected results was described in listing the objectives of operations research. The mode of explanation of a complex situation usually takes the form of classifications, characterizations, structures, and models.

The first three forms of results are most frequently employed in describing phenomena; for example, historical demands for a product, or setting down in understandable form large masses of information, for example, the relation of parts and subassemblies to a line of finished products

The fourth form, the model is usually employed as a systems representation of the relationships of phenomena to each other, their interactions on each other, and with decisions made by people in the real situation. The model displays the effect of these interactions or the changes in them, upon some desired total outcome or the actions needed to achieve theoretically some desired

sciences. During this period the

disciplines, were most frequently employed to obtain a balanced approach approximating that of a comprehensive discipline. Since then operations research has been extensively applied by business as well as by the military.

In the nonmilitary field, four types of activities are carried on: (1) development of special-purpose techniques such as queueing theory, allocation theory, or replacement theory, (2) application of the special-purpose techniques to specific situations in business, (3) study of opportunities for and organizational placement of operations research to realize these opportunities, and (4) elaboration of the discipline and its subject matter to learn the valid and most effective basis for its use in an organization having social and economic objectives.

Currently, greatest attention is to applications of special-purpose techniques. These applications are largely as adjuncts to functions already in existence in business organizations such as industrial engineering. Operations research is gradually being recognized by business managers as a separate profession having unique and useful purposes.

This slowness may be a natural development because of (1) the instinctive desire to do small-scale testing first, (2) the longer-term commitment of people and resources needed for the development of broader areas of understanding, and (3) the inevitable existence of problems which themselves are symptoms of larger challenges, but which are, nevertheless, tractable to the methods of operations research

These applications have the desirable result in showing the applicability and effectiveness of this approach. However, there is the danger that the larger implications for the work may be lost in the process. As an example, a frequent application is in the allocation of effort and resources for optimal production scheduling, applying the technique of linear programming. In this situation there are usually sufficient elements that the relationship among them is not intuitively discerned, or the nature of the actions to be taken to obtain an optimal solution is not apparent.

To use the particular technique, it is usually assumed that (1) what is to be produced in a given period is known and can be held fixed, (2) judgment of merit is known and can be defined as a single-valued mathematical function, (3) the relation of volume to cost is known and may be assumed as linear, at least over a known range, and (4) the production process can be adequately described by a set of mathematical equations (most usually linear) that tell how much of each resource involved in the process goes into each product What is commonly sought is a specific answer to an aggravating problem.

In this approach, the same tools are used that are being used in operations research But in its purpose and basic assumptions, as well as in its results, this approach differs sharply from operations research and should not be confused with it. In the first place, it assumes the reality of the symptom, whereas a basic assumption of operations research is that a symptom only indicates where work is needed but does not, by itself, indicate the nature of the problem. Secondly, this common approach assumes that problems in a business or in other risk-taking organizations can be treated in isolation, whereas operations research always assumes that its subject matter is an interrelated and interdependent system having problems which can only be treated effectively as systemic rather than mechanic.

The distinction between problems treated within such limiting conditions and problems treated as symptomatic of underlying causes within a system may well delineate the approximate dividing line between the use of operations research for efficiency purposes, as in industrial engineering directed toward obtaining most efficient solutions of the designs under stated conditions (where the statements of conditions are obtained by other means and from other sources), and operations research as a unique technique directed toward obtaining and disclosing more explicit understanding.

The increasing size and diversity of business organizations, the increasing scope and rapidity of technological change, and the increasing social expectation of society from such business organizations pose new challenges and greater opportunities. The need for more explicit understanding of these new situations and for appropriate information with which to continue to achieve effective joint performance from business organizations creates the opportunities for operations research

It is from a perception of the new demands that impetus to the development of operations research is likely to come. While the tools of operations research in their origin were not designed for problems of this kind but rather, as the name implies, for the analysis of the complexities in immediate operations, the people who sense and anticipate these new basic needs for understanding increasingly see in operations research a promising approach for the identification, description, and understanding of the situations created by these needs and demands. *See* INDUSTRIAL ENGINEERING.

[M L H]

Bibliography: J. F. McCloskey and F. N Trefethen (eds), *Operations Research for Management*, vol. 1, 1954; Operations Research Group at Case Institute, *Comprehensive Bibliography on Operations Research*, 1958

Operator theory

At one level of abstraction an operator is simply a function whose arguments and values are real- (or complex-) valued functions of one or more real variables; in more naive terms an operator is a rule for converting such real (or complex-) valued functions into others The following are simple examples: (1) the operator which takes each differentiable real-valued function of one variable into its derivative; (2) the operator which takes each twice-differentiable function f of one variable into

$$\left(\frac{df}{dx}\right)^2 + 2\frac{d^2f}{}$$

(3) the operator which takes each twice-differentiable function f of three variables into

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

, the operator which takes the continuous func-
n f of one real variable into the function g where

$$g(x) = \int_0^1 \sqrt{x+y}\, f(y)\, dy$$

Since an operator is a function, the usual func-
nal notation is applicable. $L(f)$ may be used to
note the result of operating on f with the opera-
r L. The set of all functions f for which $L(f)$ is
fined is called the domain of L and the set of all
nctions g such that $L(f) = g$ for some f in the
omain of L is called the range of L. It is obvious
at solving a differential or integral equation is
quivalent (in many ways) to solving an operator
quation $L(f) = g$ where g and L are given and it
s required to find f. Moreover the operator concept
an be very useful both in theory and practice, pro-
ducing a great variety of illuminating insights.

In large part the fruitfulness of the operator concept can be traced to two sources. One of these is the possibility of adding and multiplying operators in such a way that many, though not all, of the laws of ordinary algebra hold. The other is the fact that the ranges and domains of operators behave in many respects like ordinary space and indeed may be regarded as contained in infinite dimensional generalizations of the familiar three-dimensional space of solid geometry. This makes it possible to think of an operator as a geometrical transformation and to exploit one's spatial intuition.

Let D be a family of real-valued functions such that $\lambda f + \mu g$ is in D whenever λ and μ are real numbers and f and g are in D. Let L and M be operators with domain D and range included in D. Then the operator which takes each f in D into $L[M(f)]$ is called the product of L and M and is denoted by LM. Moreover the operator which takes each f in D into $L(f) + M(f)$ is called the sum of L and M and denoted by $L + M$. In particular one may form powers L^2, L^3, . . . , polynomials $a_0 + a_1L + a_2L^2 + \cdots + a_rL^r$, and in suitably restricted contexts power series. It is important to note that while it is always true that $L + M = M + L$ it is not always true that $LM = ML$. On the other hand it is possible to show that $(LM)N = L(MN)$ and that $(L + M) + N = L + (M + N)$ for all L, M, and N so that parentheses may be omitted just as in ordinary algebra.

In many but not all cases the sets of functions with which one deals derive their spacelike properties from the possibility of assigning a distance $\rho(f,g)$ to each pair f and g of members of the set D under consideration. This is done in such a manner that $\rho(f,g) = \rho(g,f)$, $\rho(f,g) > 0$ if $f \neq g$; $\rho(f,f) = 0$ and $\rho(f,g) \leqq \rho(f,h) + \rho(h,g)$ for all f, g, and h in D. When D is as described in the preceding paragraph, ρ is often chosen so that $\rho(f,g) = \|f - g\|$ where $\|f\| = \rho(f,0)$. If $\|\lambda f\| = |\lambda|\,\|f\|$ for all real numbers λ then $\|f\|$ is said to be a norm for D. There will usually be more than one way of norming a given D. For example, if D is the set of all real-valued continuous functions defined on the interval $0 \leqq x \leqq 1$, setting

$$\|f\| = \max_{0 \leq x \leq 1} |f(x)|$$

gives one value for D, and setting

$$\|f\|_1 = \sqrt{\int_0^1 |f(x)|^2\, dx}$$

gives another value for D. The analogy with the familiar space of experience is closest when the [illegible]

f and g are in the domain of L, and λ and μ are numbers. Insofar as there is a general theory of operators, it is largely concerned with linear operators, and this article will discuss linear operators exclusively. It is useful to develop this theory from axioms.

Axioms. Let F denote either the field of all real numbers or the field of all complex numbers. A vector space over F is a set or collection X whose members are of an unspecified character except that they may be added together and multiplied by the members of F in such a way that the following formal laws are satisfied:

1. $(f + g) + h = f + (g + h)$ and $f + g = g + f$ for all f, g, and h in X.
2. There is a unique zero vector 0 in X such that $f + 0 = f$ for all f in X.
3. $\lambda(\mu f) = (\lambda\mu)f$, $(\lambda + \mu)f = \lambda f + \mu f$, and $\lambda(f + g) = \lambda f + \lambda g$ for all f and g in X and all λ and μ in F.
4. $1f = f$ for all f in X.

By generalizing the more special and concrete definition in the obvious fashion a linear operator is defined to be a function L whose domain is a vector space X, whose range is in a vector space Y, and for which it is true that $L(\lambda f + \mu g) = \lambda L(f) + \mu L(g)$ whenever f and g are in X and λ and μ are in F.

Finite dimensional case. A vector space X is said to be finite dimensional if it contains a finite subset $v_1, v_2, \ldots, v_n$ spanning the space in the sense that every element in the space may be written in the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$ where the λ_j are in F. The representation $f = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$ is unique if and only if no v_j is in the span of the rest. In this case $v_1, v_2, \ldots, v_n$ is said to form a basis for X and the λ_j are said to be the coordinates of f with respect to this basis. It is not hard to show that any two bases for the same space have the same number of elements. This number is called the dimension of the space. Let L be a linear operator whose domain X is finite dimensional. It follows immediately that the range is also finite dimensional and that the dimension d_R of the range is less than or equal to the dimension d_X of the do

In the nonmilitary field, four types of activities are carried on: (1) development of special-purpose techniques such as queueing theory, allocation theory, or replacement theory, (2) application of the special-purpose techniques to specific situations in business, (3) study of opportunities for and organizational placement of operations research to realize these opportunities, and (4) elaboration of the discipline and its subject matter to learn the valid and most effective basis for its use in an organization having social and economic objectives

Currently, greatest attention is to applications of special-purpose techniques These applications are largely as adjuncts to functions already in existence in business organizations such as industrial engineering. Operations research is gradually being recognized by business managers as a separate profession having unique and useful purposes.

This slowness may be a natural development because of (1) the instinctive desire to do small-scale testing first, (2) the longer-term commitment of people and resources needed for the development of broader areas of understanding, and (3) the inevitable existence of problems which themselves are symptoms of larger challenges, but which are, nevertheless, tractable to the methods of operations research

These applications have the desirable result in showing the applicability and effectiveness of this approach. However, there is the danger that the larger implications for the work may be lost in the process As an example, a frequent application is in the allocation of effort and resources for optimal production scheduling, applying the technique of linear programming. In this situation there are usually sufficient elements that the relationship among them is not intuitively discerned, or the nature of the actions to be taken to obtain an optimal solution is not apparent

To use the particular technique, it is usually assumed that (1) what is to be produced in a given period is known and can be held fixed, (2) judgment of merit is known and can be defined as a single-valued mathematical function, (3) the relation of volume to cost is known and may be assumed as linear, at least over a known range, and (4) the production process can be adequately described by a set of mathematical equations (most usually linear) that tell how much of each resource involved in the process goes into each product. What is commonly sought is a specific answer to an aggravating problem.

In this approach, the same tools are used that are being used in operations research. But in its purpose and basic assumptions, as well as in its results, this approach differs sharply from operations research and should not be confused with it In the first place, it assumes the reality of the symptom, whereas a basic assumption of operations research is that a symptom only indicates where work is needed but does not, by itself, indicate the nature of the problem Secondly, this common approach assumes that problems in a business or in other risk-taking organizations can be treated in isolation, whereas operations research always assumes that its subject matter is an interrelated and interdependent system having problems which can only be treated effectively as systemic rather than mechanic.

The distinction between problems treated within such limiting conditions and problems treated as symptomatic of underlying causes within a system may well delineate the approximate dividing line between the use of operations research for efficiency purposes, as in industrial engineering directed toward obtaining most efficient solutions of the designs under stated conditions (where the statements of conditions are obtained by other means and from other sources), and operations research as a unique technique directed toward obtaining and disclosing more explicit understanding.

The increasing size and diversity of business organizations, the increasing scope and rapidity of technological change, and the increasing social expectation of society from such business organizations pose new challenges and greater opportunities. The need for more explicit understanding of these new situations and for appropriate information with which to continue to achieve effective joint performance from business organizations, creates the opportunities for operations research

It is from a perception of the new demands that impetus to the development of operations research is likely to come. While the tools of operations research in their origin were not designed for problems of this kind but rather, as the name implies, for the analysis of the complexities in immediate operations, the people who sense and anticipate these new basic needs for understanding increasingly see in operations research a promising approach for the identification, description, and understanding of the situations created by these needs and demands. See INDUSTRIAL ENGINEERING.

[M L.H]

Bibliography: J. F McCloskey and F. N. Trefethen (eds), *Operations Research for Management*, vol. 1, 1954; Operations Research Group at Case Institute, *Comprehensive Bibliography on Operations Research*, 1958.

Operator theory

At one level of abstraction an operator is simply a function whose arguments and values are real- (or complex-) valued functions of one or more real variables; in more naive terms an operator is a rule for converting such real- (or complex-) valued functions into others The following are simple examples: (1) the operator which takes each differentiable real-valued function of one variable into its derivative; (2) the operator which takes each twice-differentiable function f of one variable into

$$\left(\frac{df}{dx}\right)^2 + x^2\frac{d^2f}{\ldots}$$

such that $L(f_1) - \lambda f_1 = g_1$; (4) N_λ is finite dimensional It follows easily that existence and uniqueness questions for the operator equation $L(f) - \lambda f = g$ reduce to the corresponding questions for the restriction of L to the finite-dimensional space N_λ and hence that the simple analysis given earlier applies. It may be proved further that M_λ coincides with X for all values of λ except those in a sequence λ_1, λ_2 . . . such that $\lim_{n\to\infty} \lambda_n = 0$.

Let K be a continuous real-valued function defined *on the unit square* $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$. Let X be the Banach space of all continuous real-valued functions defined on the interval $0 \leqq x \leqq 1$ with $\|f\| = \max_{0\leq x\leq 1} |f(x)|$. Then it can be proved that the *operator* L_K which takes f into g where

$$g(x) = \int_0^1 K(x,y)f(y)\,dy$$

is a completely continuous linear operator. Application of the theorems quoted above yields most of the results of the Fredholm theory of integral *equations* Such integral operators occur in inverting *the* members of a large class of linear *differential operators* There are similar results for integral operators in $2n$ variables. it being possible to *establish complete* continuity whenever the region of integration is bounded.

Infinite-dimensional versions of theorems about bases of proper vectors and functions of operators take their simplest and *most* complete form when *the* underlying Banach space is a Hilbert space; that is, when there is defined an F-valued "inner product" $f \cdot g$ for each f and g in X which satisfies the following conditions (1) $(\lambda f + \mu g) \cdot h = \lambda(f \cdot h) + \mu(g \cdot h)$, (2) $(f \cdot g) = \overline{(g \cdot f)}$; and (3) $(f \cdot f) = \|f\|^2$ for all f, g, and h in X and all λ and μ in F The simplest (and original) *example* of a Hilbert space is the vector space of all sequences $c_1, c_2, \ldots$ of complex numbers such that $|c_1|^2 + |c_2|^2 + \cdots < \infty$, based on the definition $(c_1, c_2, \ldots) \cdot (c_1', c_2', \ldots) = c_1\overline{c_1'} + c_2\overline{c_2'} + \cdots$

Now let L be a completely continuous linear operator the domain of which is a separable Hilbert space H and whose range is contained in H Let L be self adjoint in the sense that $L(f) \cdot g = f \cdot L(g)$ for all f and g in H Then it is a theorem that there exists a sequence v_1, v_2, *of members* of H which has the following *properties* (1) each v_j is a proper vector for L, (2) if f is any element in H and $c_j = f \cdot v_j$ then $f = c_1v_1 + c_2v_2 + \cdots$ in the sense that the partial sums of this *series have* f as a limit; (3) $v_i \cdot v_j = 0$ if $i \neq j$ and $v_i \cdot v_i = 1$ Such a sequence is said to be an orthonormal basis for H. Just as in the finite dimensional case *one* can form more or less arbitrary functions of the operator L.

In rough *terms* the celebrated *spectral* theorem is a generalization of the preceding theorem *in* which the operator L, though self adjoint, *is* not required to be completely continuous Instead of a discrete basis of proper vectors one finds a sort of continuous basis. More precisely *it* is possible to map H onto a Hilbert space H' whose elements are complex-valued functions in such a *manner that* the norms and the vector space operations are preserved and so that L *becomes the operator of multi*plying the elements of H' by a fixed real-valued function.

In addition to the abstract theory there are many detailed studies of particular operators of importance such as the Laplace and Fourier *transforms* and numerous applications to differential and integral equations Moreover in addition *to the theory* of single operators there are extensive theories of *certain kinds of collections* of operators. A ring of operators contains the sum, product, and difference of any two of its members, and a *group of* operators contains the inverse of every operator in it and the product of each two operators in it The theory of groups and rings of operators is related to the algebraic theory of groups *and rings much* as the theory of a single operator is related to *systems of* linear algebraic equations, *It has applications* to harmonic analysis and to the conceptual *foundations* of quantum mechanics *See* COMPLEX NUMBERS AND COMPLEX VARIABLES; GROUP THEORY; INTEGRAL TRANSFORM; REAL VARIABLE; RING THEORY, SET THEORY; TOPOLOGY. [C.W.M]

Bibliography: A E Taylor, *Introduction to Functional Analysis*, 1958

Operator training

Teaching a factory or office worker to perform a job Instruction may be limited to the performance of one repetitive operation, *or it* may cover all skills *required* of a craftsman *such* as a pipefitter or toolmaker.

Basically, operator training is the *same* as *other* training In practice, operator *training* usually has more limited objectives and scope *than* training given in most schools. Factories and offices *normally* train an operator to meet *only* their own specific needs

Operator training is done in two *ways*. entirely on the job. or by planned instruction. On-the-job training is ...
lear
stan
the ... first cost it usually produces long learning *periods*, low output, poor quality, *misuse* of tools and materials, and unsafe working habits Planned *training* by a competent *instructor*, followed by guided, on-the-*job* practice has a relatively high initial cost but pays off *rapidly* in *enhanced* operator performance

Most courses are prepared by the training department, by the *industrial engineering* department of the company, or by *an* outside specialist *Small* companies may appoint a *supervisor* or other person to organize a course *in* addition to his *other* duties Instruction is usually given by a supervisor, experienced operator, or other person of recognized competence Keys to good results are (1) to get participation *from* the *trainee* and the trainee's supervisor from the outset *and* (2) to have the in-

main Let Y be the vector space containing the range of L and let d_Y denote the dimension of Y. This gives $d_R \leqq d_Y$ and $d_R \leqq d_X$. The differences $d_Y - d_R$ and $d_X - d_R$ measure the extent to which the operator equation $L(f) = g$ fails to have a unique solution for all g. In fact if X_1 and X_2 are finite-dimensional subvector spaces of a vector space and $X_1 \subseteq X_2$ then $X_1 = X_2$ if and only if X_1 and X_2 have the same dimension. Thus $L(f) = g$ is always solvable if and only if $d_Y = d_R$ and $d_Y - d_R$ is a measure of the size of the set of gs for which no solution exists. On the other hand it is easily seen that if f_0 is a particular solution of $L(f) = g$ then the general solution is $f_0 + h$ where h is any element of X such that $L(h) = 0$. The set of all such h is a vector space N, called the null space of L, whose dimension d_N measures the extent to which the equation $L(f) = g$ has multiple solutions. It is not hard to show that $d_N + d_R = d_X$ so that $d_N = d_X - d_R$. *In the special but important case in which* $X = Y$, $d_X - d_R = d_Y - d_R$. Thus either $L(f) = g$ has a unique solution for all g (nonsingular case) or else for many values of g, $L(f) = g$ has no solutions, and whenever it has any nonzero solutions it has many (singular case). There is a certain sense in which most operators are nonsingular. Let I denote the identity operator *which takes every vector into itself*. Then it can be shown that $L - \lambda I$ is nonsingular for all but a finite number of values of λ Indeed $L - \lambda I$ will be singular if and only if there exists a nonzero vector f such that $L(f) - \lambda f = 0$, that is, $L(f) = \lambda f$. Such an f is called a proper vector (or eigenvector) belonging to the proper value (or eigenvalue) λ It is easy to show that proper vectors belong*ing to distinct proper values* are linearly independent in the sense that no one is in the span of the rest Thus the number of distinct proper values cannot exceed the dimension of the space. *See* DIFFERENTIAL EQUATION

Knowledge of the proper values of an operator L yields a great deal of information about the nature of L, especially in the important case in which the *domain X admits a basis made up of proper* vectors. Let $v_1, v_2, \ldots, v_n$ be a basis for X and let $L(v_j) = \lambda_j v_j$ where the λ_j are (not necessarily distinct) members of F Very simple computations lead to the following observations

1. Every proper value of L is equal to some λ_j
2. L is nonsingular if and only if no λ_j is zero
3. If L is nonsingular, the inverse operator carries $\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n$ into $(\mu_1/\lambda_1)v_1 + (\mu_2/\lambda_2)v_2 + \cdots + (\mu_n/\lambda_n)v_n$
4. If P is any polynomial with coefficients in F then $P(L)$ carries $\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n$ into $P(\mu_1)v_1 + P(\mu_2)v_2 + \cdots + P(\mu_n)v_n$.

The structure of $P(L)$ revealed by (4) suggests a definition of $F(L)$ where F instead of being a polynomial is an arbitrary function with domain and range in F. To wit: $F(L)(\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n) = F(\mu_1)v_1 + F(\mu_1)v_2 + \cdots + F(\mu_n)v_n$. A similar, but of course more subtle, definition in certain infinite-dimensional cases is the source of the modern rigorization of the celebrated operational calculus of O Heaviside.

Returning to the case in which X need not equal Y, let $v_1, v_2, \ldots, v_n$ and $w_1, w_2, \ldots, w_m$ be bases for X and Y respectively. Let $L(v_j) = \alpha_{j1}w_1 + \alpha_{j2}w_2 + \cdots + \alpha_{jn}w_n$ where each α_{ji} is in F Then $L(x_1 v_1 + \cdots + x_n v_n) = x_1 L(v_1) + x_2 L(v_2) + \cdots + x_n L(v_n) = y_1 w_1 + y_2 w_2 + \cdots + y_m w_m$ where $y_i = \alpha_{1i}x_1 + \alpha_{2i}x_2 + \ldots \alpha_{ni}x_n$. The rectangular array in which α_{ji} is in the ith row and jth column is called the matrix of L with respect to the basis in question. It is clear that solving the operator equation $L(f) = g$ in the finite-dimensional case is equivalent to solving m linear algebraic equations in n unknowns The theory sketched above is the theory of such equations couched in the language of operator theory. When $X = Y$ and $v_j = w_j$ for all j then the condition $\alpha_{ji} = \bar{\alpha}_{ij}$ (where the overbar denotes complex conjugate) implies *that there exists a basis for X made* up of proper vectors of L. However this condition is by no means a necessary one.

Infinite dimensional case. When X is not assumed to be finite dimensional, such simple and general theorems as those described above are no longer available. In certain contexts, however, more complicated and less complete analogs of them may be proved It is with these *that the* general theory of linear operators is mainly concerned

Let X be a vector space which is not necessarily finite dimensional but instead is equipped with a norm—defined in the abstract case as suggested by the definition given above for real functions spaces. Such a normed vector space is said to be complete if each sequence $f_1, f_2, \ldots$ of members of X which is convergent in the sense that

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \|f_n - f_m\| = 0$$

is also convergent in the sense that

$$\lim_{n\to\infty} \|f_n - f\| = 0$$

for some f in X This f is easily seen to be *unique* and is called the limit of the sequence $\{f_n\}$. A complete normed vector space is called a Banach space A normed vector space X is said to be separable if there exists a sequence $f_1, f_2, \ldots, f_n$ of elements of X such that every element of X is the limit of some subsequence. The space of continuous functions defined earlier is a separable Banach space. Now let $X = Y$ and let X be a Banach space *Let L be completely* continuous in the sense that whenever $f_1, f_2, \ldots$ is a sequence of elements such that $\|f_n\| \leqq 1$ for all n there is a subsequence $f_{n_1}, f_{n_2}, \ldots$ such that $L(f_{n_1}), L(f_{n_2}), \ldots$ is convergent in the first of the two senses defined above. Then the following theorem can be proved. For each λ in F with $\lambda \neq 0$ there is a pair of vector subspaces M_λ and N_λ of X such that: (1) $L(f)$ is in M_λ for all f in M_λ and $L(f)$ is in N_λ for all f in N_λ; (2) every f in X can be written uniquely in the form $f_1 + f_2$ where $f_1 \in M_\lambda$ and $f_2 \in N_\lambda$; (3) for each $g_1 \in M_\lambda$ there is one and only one element $f_1 \in M_\lambda$

such that $L(f_1) - \lambda f_1 = g_1$; (4) N_λ is finite dimensional. It follows easily that existence and uniqueness questions for the operator equation $L(f) - \lambda f = g$ reduce to the corresponding questions for the restriction of L to the finite-dimensional space N_λ and hence that the simple analysis given earlier applies. It may be proved further that M_λ coincides with X for all values of λ except those in a sequence $\lambda_1, \lambda_2, \ldots$ such that $\lim_{n\to\infty} \lambda_n = 0$.

Let K be a continuous real-valued function defined on the unit square $0 \leqq x \leqq 1, 0 \leqq y \leqq 1$. Let X be the Banach space of all continuous real-valued functions defined on the interval $0 \leqq x \leqq 1$ with $\|f\| = \max_{0\leq x\leq 1} |f(x)|$. Then it can be proved that the operator L_K which takes f into g where

$$g(x) = \int_0^1 K(x,y)f(y)\,dy$$

is a completely continuous linear operator. Application of the theorems quoted above yields most of the results of the Fredholm theory of integral equations. Such integral operators occur in inverting the members of a large class of linear differential operators. There are similar results for integral operators in $2n$ variables, it being possible to establish complete continuity whenever the region of integration is bounded.

Infinite-dimensional versions of theorems about bases of proper vectors and functions of operators take their simplest and most complete form when the underlying Banach space is a Hilbert space: that is, when there is defined an F-valued "inner product" $f \cdot g$ for each f and g in X which satisfies the following conditions: (1) $(\lambda f + \mu g) \cdot h = \lambda(f \cdot h) + \mu(g \cdot h)$; (2) $(f \cdot g) = \overline{(g \cdot f)}$, and (3) $(f \cdot f) = \|f\|^2$ for all f, g, and h in X and all λ and μ in F. The simplest (and original) example of a Hilbert space is the vector space of all sequences $c_1, c_2, \ldots$ of complex numbers such that $|c_1|^2 + |c_2|^2 + \cdots < \infty$, based on the definition $(c_1,c_2, \ldots) \cdot (c_1',c_2', \ldots) = c_1\overline{c_1'} + c_2\overline{c_2'} + \cdots$.

Now let L be a completely continuous linear operator the domain of which is a separable Hilbert space H and whose range is contained in H. Let L be self adjoint in the sense that $L(f) \cdot g = f \cdot L(g)$ for all f and g in H. Then it is a theorem that there exists a sequence $v_1, v_2, \ldots$ of members of H which has the following properties: (1) each v_i is a proper vector for L, (2) if f is any element in H and $c_i = f \cdot v_i$ then $f = c_1v_1 + c_2v_2 + \cdots$ in the sense that the partial sums of this series have f as a limit, (3) $v_i \cdot v_j = 0$ if $i \neq j$ and $v_i \cdot v_i = 1$. Such a sequence is said to be an orthonormal basis for H. Just as in the finite-dimensional case one can form more or less arbitrary functions of the operator L.

In rough terms the celebrated spectral theorem is a generalization of the preceding theorem in which the operator L, though self-adjoint, is not required to be completely continuous. Instead of a discrete basis of proper vectors one finds a sort of continuous basis. More precisely it is possible to map H onto a Hilbert space H' whose elements are complex-valued functions in such a manner that the norms and the vector space operations are preserved and so that L becomes the *operator of multiplying* the elements of H' by a fixed real-valued function.

In addition to the abstract theory there are many detailed studies of particular operators of importance such as the Laplace and Fourier transforms and numerous applications to differential and integral equations. Moreover in addition to the theory of *single* operators there are extensive theories of certain kinds of collections of operators. A *ring* of operators contains the sum, product, and difference of any two of its members, and a group of operators contains the inverse of every operator in it and the product of each two operators in it. The theory of groups and rings of operators is related to the algebraic theory of groups and rings much as the theory of a single operator is related to systems of linear algebraic equations. It has applications to harmonic analysis and to the conceptual foundations of quantum mechanics. *See* COMPLEX NUMBERS AND COMPLEX VARIABLES; GROUP THEORY; INTEGRAL TRANSFORM; REAL VARIABLE; RING THEORY; SET THEORY; TOPOLOGY. [G.W.M.]

Bibliography: A. E. Taylor, *Introduction to Functional Analysis*, 1958.

Operator training

Teaching a factory or office worker to perform a job. Instruction may be limited to the performance of one repetitive operation, or it may cover all skills required of a craftsman such as a pipefitter or toolmaker.

Basically, operator training is the same as other training. In practice, operator training usually has more limited objectives and scope than training given in most schools. Factories and offices normally train an operator to meet only their own specific needs.

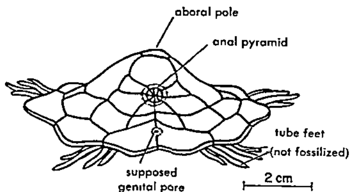
Operator training is done in two ways: entirely on the job, or by planned instruction. On-the-job training is comparatively unplanned. The operator learns from the foreman or group leader as circumstances of the job permit. While this training has the advantage of low first cost, it usually produces long learning periods, low output, poor quality, misuse of tools and materials, and unsafe working habits. Planned training by a competent instructor, followed by guided, on-the-job practice has a relatively high initial cost but pays off rapidly in enhanced operator performance.

Most courses are prepared by the training department, by the industrial engineering department of the company, or by an outside specialist. Small companies may appoint a supervisor or other person to organize a course in addition to his other duties. Instruction is usually given by a supervisor, experienced operator, or other person of recognized competence. Keys to good results are (1) to get participation from the trainee and the trainee's supervisor from the outset and (2) to have the in-

struction done by experienced authorities on the subject Operator training emphasizes the building of confident, adequate people rather than merely the training of technically competent workers. This tends to create a desire to do the job properly in addition to the ability to do it. *See* METHODS ENGINEERING. [H.T.S.]

Ophiocystioidea

A small class of extinct Eleutherozoa which ranged from Ordovician to Devonian times (see illustration). The domed aboral surface of the test was roofed by polygonal plates and carried an anal pyramid. Below, at the center of the flat adoral side lay the mouth, with 5 interradial jaws surrounded by a flexible peristome. From it radiated the 5 ambulacra, each one with 3 rows of ambulacral plates. About 8 pairs of large tube-feet emerged from each ambulacrum, by way of pores

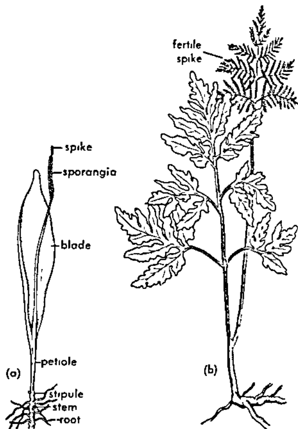


Volchovia volborthi (Ordovician, Russia).

lying between the median and the outer rows of plates. Each tube-foot was coated with small overlapping scales. In each of the 5 adoral interradial areas lay a single row of plates, and in one of them was the madreporite *See* ELEUTHEROZOA [H.B F.]

Ophioglossales

An order of the class Filicineae commonly known as the adder's tongue ferns. It is a small group with only 3 genera and about 80 species. Two of the genera, *Ophioglossum* and *Botrychium*, are widely distributed in tropical and temperate regions and have about the same number of species, while the third genus, *Helminthostachys*, is represented by a single species confined to southeastern Asia and Polynesia. These are considered the most primitive of the present-day ferns. No fossils have been reported for this order. The plants are homosporous and eusporangiate, that is, spore sacs develop from groups of epidermal cells. This group is distinguished from other ferns by the arrangement of the sporogenous tissue in the characteristic "fertile spike" of the sporophyte (spore-producing generation). The leaves are erect and not circinate *See* LEAF (BOTANY). The gametophyte (gamete-producing generation) is a small colorless, fleshy, subterranean saprophyte (living on dead organic



(a) *Ophioglossum*, or adder's tongue fern, with a single leaf, consisting of a petiole and simple, oval blade, and a stalked, fertile spike with several large sporangia (b) *Botrychium*, or grape fern, showing the division of the leaf into a sterile foliage part and a fertile spike with many sporangia. (*From H J Fuller and O Tippo, College Botany, Holt, rev. ed, 1954*)

matter) *associated with an* endophytic fungus (*see* FUNGI) The order appears to be an evolutionary dead-end *See* FILICINEAE; PTEROPSIDA. [P.A.]

Bibliography: *See* FILICALES.

Ophiurae

A suborder of Ophiurida in which the vertebrae articulate by means of ball-and-socket joints, and the arms, which do not branch, move mainly from side to side and do not coil in the vertical plane The disk and arms are usually sheathed in regularly arranged plates These are disposed in four series on the arms, namely one dorsal, one ventral, and two lateral There is a single madreporite. The suborder embraces most of the known genera of brittle stars and includes 13 families. *See* EURYALAE, OPHIURIDA [H B F]

Ophiurida

The largest order of Ophiuroidea The order includes all extant species and known fossil forms from the Triassic onward, as well as one Paleozoic family, the Aganasteridae The ambulacral plates are fused together in pairs to form vertebrae, and the ambulacral groove has closed over to become a purely internal canal. The order comprises two suborders. *See* EURYALAE; OPHIURAE; OPHIUROIDEA. [H B F.]

Ophiuroidea

A subclass of the Asterozoa, known as the brittle stars, in which the arms are usually clearly demarcated from a central disk and perform whiplike locomotor movements, and the tube-feet are nonsuctorial sensory tentacles. In all existing ophiuroids the ambulacral plates fuse together in pairs to form articulating joints termed vertebrae, and the ambulacral groove is converted into an internal epineural canal

As Fig 1 shows, typical ophiuroids have a distinctive shape unlike that of the other Asterozoa, and the vigorous lashing movements of the arms distinguish them at once from asteroids However, fossil ophiuroids are known which approach asteroids in structure, and the two groups are evidently closely related. See ASTEROIDEA; ASTEROZOA.

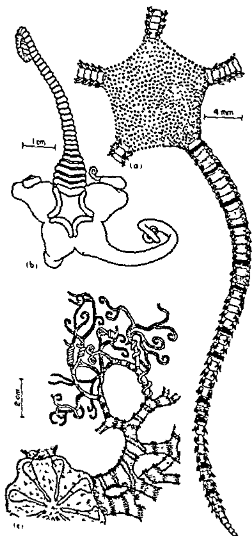


Fig 1 Representative ophiuroids. (a) *Pectinura cylindrica* (b) *Astroporpa wilsoni*. (c) *Gorgonocephalus chilensis*

There are about 1900 extant species referred to 220 genera, arranged to form 1 order (see OPHIURIDA). Within it 2 suborders are recognized, the Ophiurae and Euryalae. There are also 2 Paleozoic orders, the Stenurida and Oegophiurida

Ophiuroids are usually 5-armed, with 4- or 6-armed individuals occurring as abnormalities; but a few species are regularly 6- or 7-armed. In some Euryalae the arms may branch repeatedly; these are the so-called basket fishes. In the smallest species the disk and arms together may be no more than a few millimeters across Large species may have a disk 10 cm across and an armspread of 50 cm. Tropical species are often patterned in contrasting colors but most ophiuroids tend to match their environment. Their biochromes do not include echinochromes. Some species are luminescent, although not constantly so; the light is emitted by cells at the bases of the arm spines, usually only on stimulation. Such species are known for the genera *Amphiura*, *Amphipholis*, and *Ophiacantha*.

Relation to man. No ophiuroids are used as food by man, and none are venomous. Ophiuroids must have a considerable indirect economic importance in view of their immense numbers and consequent significance in natural food chains involving commercially sought species.

Ecology. Ophiuroids occur in all the oceans from low-tide level downward, often in dense populations which number millions to the hectare. Six families range below a depth of 2 miles, the genera *Ophiura*, *Amphiophiura*, and *Ophiacantha* range below 4 miles. The shallow-water forms hide among algae, under stones, or within sponges, or bury the disk in sand or mud leaving only the arms protruding Deep-water forms lie in or on the bottom material, or adhere to corals or cidarids.

Among their numerous parasites are Protozoa in the stomach or genital organs; nematodes, trematodes, and Crustacea, Myzostomida sometimes occur (see CRINOIDEA) Parasitic algae such as *Coccomyxa ophiurae* infest the spines and cause malformations; parasitic mollusks are much [illegible]

[illegible] carries a median row of plates In five Paleozoic families the ambulacral plates are paired, as in asteroids, in all other ophiuroids these plates fuse in pairs to produce jointed vertebrae, and the corresponding adambulacral plates become the so-called lateral plates on either side of the arm. In most extant forms the arm comprises a series of jointed segments, each one containing a vertebra, and each one covered externally by the ventral plate, right and left lateral plates, and a dorsal plate. These features are used in diagnosing the several orders (see Oegophiurida, Ophiurida, Stenurida) Spines are often carried by the skeletal plates, especially where the lateral arm plates represent the adambulacral spines of asteroids In the sub-

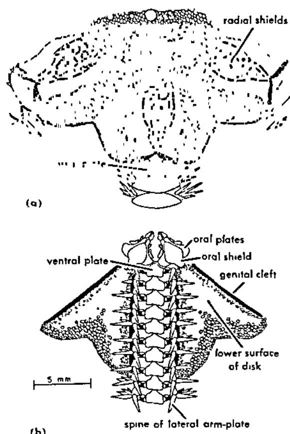


Fig. 2. External features of the ophiuroid *Amphiura* (a) Aboral. (b) Adoral.

order Ophiurae they commonly form an erect fringe to the sides of the arms; in the suborder Euryalae they commonly are transformed into clubs or hooklets, and hang downward

Muscular system. The arms in extant ophiuroids are provided with well-developed longitudinal muscles linking the successive vertebrae. The two extant suborders are each characterized by a distinctive arm movement, horizontal in the Ophiurae, and vertical in the Euryalae This distinction is related in either case to the form of the vertebrae, which in turn are defined in the taxonomic diagnoses. Ophiurae move rapidly when disturbed. One of the arms is either trailed or pushed ahead, whereas the other four arms operate as two opposite pairs of levers, thrusting the body forward in a series of rapid jerks. Unlike asteroids, the tube-feet play little part in locomotion, or none at all. In very young stages, however, the tube-feet may be used as stilts, and seem to be weakly adhesive. Euryalae have bigger vertebrae and smaller muscles, and their movements are less spasmodic, but they are able to coil the arms very firmly around other objects, and retain their hold after death.

Alimentary system. The mouth, in the middle of the underside of the disk, serves for both ingestion and egestion of food, because there is no anus in existing forms. The gut comprises only the saclike stomach, in the walls of which are glandular hepatic cells. The stomach sends blind caeca into the arms in the Ophiocanopidae, a character recalling the asteroids.

Although many ophiuroids are predatory on small organisms when opportunity offers, most of them evidently spend most of their time scavenging in detritus, eating whatever they find They are selective, because their gut does not permit the gross mud swallowing practiced by asteroids in similar circumstances. Thus, *Pectinura* will feed selectively on beech pollen in season in the New Zealand fiords, where the trees overhang the water Among the more consistently predatory ophiuroids are certain Euryalae which cling to the branches of black corals and browse upon the polyps Ophiuroids which live in large, dense populations evidently rely upon a steady flow of suspended matter, and there is evidence that sea-floor currents supply this, according to H. Vevers (1952) and D Hurley (1959). Basket fishes sometimes live in these conditions rhythmically sweeping the branched arms toward the mouth (H. L. Clark, 1946).

Water-vascular system. This system is typical for the phylum, but the tube-feet lack ampullae and suckers, and can be retracted into so-called tentacle pores. The madreporite normally lies in one interradius of the adoral surface, but some Euryalae have a madreporite in each interradius.

Nervous system. This system is also typical for the phylum. No organs of special sense are known, and it may be inferred that photoreceptors and chemoreceptors are distributed on the ectoderm of the ambulacral tube-feet. *See* NERVOUS SYSTEM (INVERTEBRATE).

Reproduction. Ophiuroids may take up to 2 years to reach sexual maturity, and full growth may require 3 or 4 years. The life span is unknown but may be estimated at about 5 years. Large Euryalae, such as *Gorgonocephalus*, may well live much longer. The sexes are usually separate, but a few species such as *Amphipholis squamata* are hermaphroditic In a few species the female carries a dwarf male clinging to the disk. The ovaries and testes are confined to the disk, and open indirectly to the exterior by way of interradial pouches in the integument, termed genital bursae. In the Ophiocanopidae, however, the gonads are serially paired in the basal arm joints, and do not open into bursae Those species producing ophiopluteus larvae are apparently fewer than those with direct development In viviparous species the young are retained in the bursae Regeneration of lost organs is widespread Autotomy (shedding of the arms) is practiced by most species when interfered with; the disk alone regenerates the lost members. The Amphiuridae can also regenerate the gut and gonads, which may be cast off in autotomy. Species of Ophiactidae regularly reproduce by transverse fission, so that the arms (usually 6 in such forms) occur in 2 sets, of 3 large arms and 3 small arms. In no case have discarded arms been observed to regenerate. *See* EURYALAE; OPHIURAE.

[H B F]

Bibliography: T. Lyman, Report on the Ophiuroidea, Rept. Scient. Results Voyage *Challenger*, *Zool.*, vol 5, 1883

Opiates

Drugs derived from opium, the dried juice of the Oriental poppy seed The active principles are alkaloids; three are in common use, morphine, codeine, and papaverine. Earlier crude extracts of opium, like laudanum, have been largely replaced by newer, more predictable, synthetic compounds. See ALKALOID

Opium and many of its derivatives are analgesics; that is, they relieve pain; morphine is the best present-day example It achieves its effect by depression of the cerebral cortex so that sensation is dulled, anxiety is lessened, and a somnolent state is induced Other effects include depression of vital centers, reduction of gastrointestinal activity, and production of a mildly euphoric state These have therapeutic uses in certain conditions *See* PAIN, DEEP, SENSATION.

Codeine has an effect similar to but less powerful than morphine and is used for relief of moderate pain. It also has a rather selective suppressive effect on the cough center of the brain stem so it is a common ingredient in cough medicines

Papaverine is primarily an antispasmodic which acts to relieve unwanted contraction of smooth muscles, particularly of the gastrointestinal tract, bronchi, and blood vessels

The most undesirable side effect of continued use of opiates is addiction. Immediate side effects which have dangerous possibilities are depression of respiratory centers, the decrease of body temperature, and a decrease in motor coordination *See* PAPAVERALES; POPPY.

[E G ST]

Opisthobranchia

A subclass in the class Gastropoda containing *Aplysia*, or sea hares the pteropods or sea butterflies, and the nudibranchs or sea slugs

Respiration is usually by means of gills which, when present, are posterior to the heart Respiration can be maintained however, by the external surface of the body. The visceral loop is not twisted into a figure 8 and the nervous system is concentrated around the esophagus A shell, when present, is generally small and may be external or internal. There are usually two pairs of tentacles, the operculum is absent except in the families Actaeonidae and P[illegible]

[illegible] covered, internal, or even absent. There are crawling and swimming forms

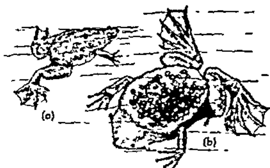
The pteropods are pelagic animals in which the foot has become modified into a pair of large fins The shells are thin and glasslike and are absent in the adults of certain families

In the nudibranchs the shell is usually absent and the gills are variable in size and in position *See* TECTIBRANCHIA, *see also* NUDIBRANCHIA

[W.J.C.]

Opisthocoela

A suborder of the Salientia in which the trunk vertebrae are convex anteriorly and concave posteriorly. Free ribs are present in the larva or adult of these frogs. The typical opisthocoelous families are the Discoglossidae with three genera, *Bombina*, *Alytes* and *Discoglossus*, in Europe, North Africa and the Orient, and a fourth, *Barbourula*, endemic to a single island in the Philippines, and the Pipidae of South America and Africa. A South American genus is *Pipa*, noted for its peculiar breeding habits The male deposits the eggs on the



Pipid toads. (a) *Xenopus mülleri*. (b) *Pipa pipa*, female with eggs (From G. K Noble, *The Biology of the Amphibia*, Dover, 1954)

back of the female. They become embedded and enclosed in the soft skin in individual pouches Protected, they develop and rupture through the skin pouches as miniature toads The important African genus is *Xenopus*, the clawed frog (see illustration), widely used in human pregnancy tests. Also included within the Opisthocoela at times, but of obscure relationship, is the monotypic Mexican family Rhynophrynidae See AMPHIBIA; SALIENTIA

[R G.Z.]

Opossum

A name given to any of several American marsupial mammals of the family Didelphidae, but usually applied to the common opossum, *Didelphis marsupialis*, of the United States and Mexico. The United States form is sometimes considered a distinct species, *D. virginiana*.



The opossum, *Didelphis virginiana*; length 21 in (From E L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Opossums occur over most of the United States east of the Rocky Mountains, and have thrived after introduction in southern California; there are also populations in the Northwest.

Opossums are prolific, tough animals that will eat anything and live anywhere, with a great capacity to survive injuries and adversity. They are among the most common fur-bearing animals and produce a beautiful silky gray fur; the pelts bring a low price, however. Opossums are hunted with dogs at night for sport and for food, and are trapped for their fur. *See* MARSUPIALIA. [J.D.B.]

Opsonin

A term used in serology and immunology to refer to a substance that enhances the phagocytosis of bacteria by leukocytes. As originally proposed by A. E. Wright and E. R. Douglas (1903–1904), the term denoted a thermolabile, relatively nonspecific substance present in normal sera. In modern usage, opsonin is more generally synonymous with the bacteriotropin of F. Neufeld and coworkers (1904–1905), a relatively thermostable antibody, increased in amount during specific immunization, that renders the corresponding bacterium more susceptible to phagocytosis. There is evidence that this action can be promoted to some extent by antibody alone, but that it is substantially increased by the further addition of the thermolabile complement system. Opsonic activity can be displayed by antibodies that also give precipitation, agglutination, lytic, and neutralization reactions

The opsonic index is a numerical measure of the opsonic activity of sera, found by dividing the average number of bacteria per phagocytic cell, as determined in the presence of an immune serum, by the corresponding value obtained in the presence of normal serum. *See* AGGLUTINATION REACTION; ANTIBODY; BACTERIA; LYTIC REACTION, NEUTRALIZATION REACTION (ANTIBODY); PHAGOCYTOSIS; PRECIPITIN TEST; SERUM (ANATOMICAL).

[H.P T.]

Bibliography: G. S. Wilson and A A. Miles, *Topley and Wilson's Principles of Bacteriology and Immunity*, 2 vols , 4th ed., 1955.

Optical activity

Optically active substances are capable of rotating the plane of polarization as plane-polarized light passes through them. There are three types of optical activity: that exhibited by substances in the crystal state only; that exhibited by substances in any physical state; and that exhibited by any substance, whatever its physical state, when placed in an intense magnetic field (Faraday effect). Since the last of these is not an intrinsic property of molecular or crystal structure alone, but must be induced by an external magnetic field, it will not be considered in this discussion. *See* FARADAY EFFECT; MAGNETOOPTICS; POLARIZED LIGHT.

Optical activity may be observed by means of a polarimeter, which is an instrument with a fixed polarizing device (nicol prism) at one end of a tube and a rotatable polarizer at the other (observer's end). If the two polarizers originally are adjusted to exclude passage of all light and an optically active substance is interposed between them, light will be observed. The angle through which the movable polarizer must be rotated in order once more to exclude the passage of light is a measure of the optical activity, the observed rotation a. Experimentally, the specific rotation $[\alpha]$ for a pure substance is given by

$$[\alpha]_\lambda^T = \frac{a}{l \cdot d}$$

wherein T is the temperature (°C), λ the wavelength of monochromatic light used (usually the sodium D line or mercury F line), l the distance through the sample (in decimeters), and d the density of the sample. When the sample is a solution of an optically active substance, d is replaced by the concentration c expressed in grams per 100 ml of solution:

$$[\alpha]_\lambda^T = \frac{a}{l \cdot c}$$

Frequently, it is of interest to compare the relative optical activities of different substances, and for this purpose the molecular rotation $[M]$ is more significant:

$$[M]_\lambda^T = [\alpha]^T \cdot \mathrm{MW}$$

wherein MW is molecular weight. Since the molecular rotation may be a very large value, it is often reported as one one-hundredth of the value obtained from the equation.

As the foregoing equations imply, optical rotatory power is a function not only of the substance, but also of the solvent (if any), the wavelength of light used, and the temperature; also, it is directly proportional to the distance the light travels through the substance or its solution.

Since plane-polarized light comprises a right and a left circularly polarized beam, any factor which retards one of these beams more than the other will effect rotation of the original plane of polarization Also, since it has been shown that the absorption coefficients of right and left circularly polarized light are different for either the dextro- or levorotatory member of a pair of enantiomorphs, provided the structure responsible for the absorption is intimately associated with the optically active center (circular dichroism, Cotton effect), it follows that the underlying source of optical activity interacts differently with a right or left circularly polarized beam Similar evidence is seen in the anomalous rotatory dispersion encountered when plane-polarized light, of wavelength appropriate to absorption by a structure intimately associated with the optically active center, is passed through an optically active substance. Furthermore, the emergent beam under such circumstances is elliptically p[illegible] indicating [illegible] change of the 90° phas[illegible] tw[illegible]

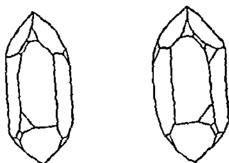


Fig 1. Left- and right-hand quartz crystals.

original right and left circularly polarized component beams

The reason for such anomalous phenomena, including optical rotation, is attributable to passage of the light through an unsymmetrical electrical field within the crystal or molecule See ROTATORY DISPERSION

Optical activity in crystals. Many inorganic crystals (*notably quartz*) and some organic crystals are *optically active* Quartz crystals are most interesting as optically active solids Well-formed crystals exhibit the phenomenon of enantiomorphism (hemihedry), that is, pairs of crystals may be found which are nonsuperposable ... of each ... levorota ... cal acti ... structure lacks a certain element of symmetry which may be described as an *n*-fold alternating axis of symmetry (where $n = 1, 2,$ or 4). Such a dissymmetric crystal structure constitutes an outwardly observable symptom of internal *electronic* dissymmetry That is, the valence electrons bonding the *elements* of the crystal are not symmetrically disposed in the crystalline aggregate Hence a beam of plane-polarized light passing through such a crystal will encounter an *electrical* field lacking in the same symmetry elements as the crystal

In a rigid system such as a crystal, a simple diagrammatic picture for optical activity is possible Figure 2 represents the variation with time of the electrical field in a beam of plane-polarized light whose d...

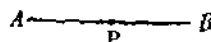


Fig 2 Plane-polarized light in a symmetrical medium

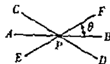


Fig 3 Plane polarized light in an unsymmetrical medium

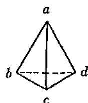
part to them a periodic motion. However, if the electrons are unsymmetrically arranged, the direction of imparted motion (polarizability) will not be along the line *AB* but along another line *CD* (Fig 3). Such a periodic motion of electrically charged particles must produce plane-polarized radiation whose direction is perpendicular to the paper, and the emergent beam will consist of two plane-polarized components of the same frequency, original (*AB*) and induced (*CD*), which combine to produce a beam whose plane of polarization falls between those of its components (*EF*). Thus the original plane of polarization has been rotated through $\theta°$ as it passed through the rigidly fixed unsymmetrical electrical field of the crystal.

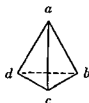
Although enantiomorphism may not be *observable in all crystalline substances*, nevertheless unsymmetrical electrical fields may exist in the crystals In enantiomorphous crystals of relatively simple inorganic substances, molecular structure and crystal structure frequently may be considered synonymous, the crystal being regarded as a macromolecule the constituent atomic groupings *of which are specifically* oriented and rigidly held by *intracrystalline forces*

Optically active crystals, comprising isotropic (sodium bromate ... *solution*, or vaporization, since the unsymmetrical field present in the crystal is lost when a change of state occurs, permitting completely random orientation of the constituent symmetrical molecules or *ions*

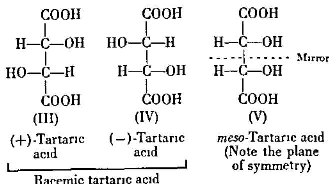
Optical activity of organic compounds. Any collection of molecules which individually possess structural dissymmetry (and therefore unsymmetrical electrical fields within them) may be treated, *on a time*-average basis, as a macromolecule *with* the symmetry properties of its components, regardless of the instantaneous orientations of the individual molecules. Accordingly liquids, vapors, and solutions of organic molecules of dissymmetrical structure will exhibit the same optical properties expected of the discrete molecules themselves Thus, as for crystals, molecules possessing an *n*-fold alternating axis of symmetry in any freely attainable conformation will fail to exhibit optical activity in the liquid, solution, or vaporized state, and those lacking such symmetry will exhibit optical activity, unless structural features permit facile interconversion of one dissymmetric conformation into its mirror image, either with or without passing through a symmetrical intermediate conformation For purposes of the ensuing discussion, only *n*-fold alternating axes of symmetry, where $n = 1$ or 2 (plane or center of symmetry), will be considered

The simplest criterion for molecular dissymmetry is the presence in a molecule of an asymmetric carbon atom, that is, a carbon atom carrying four different substituents (I) and (II). Where only one such asymmetric center is present, two nonsuper-

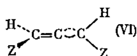




imposable mirror-image configurations are possible, one rotating the plane of polarized light to the right and the other equally to the left. However, when more than one such center is present, all configurations are not necessarily nonsuperimposable on their mirror images, for if pairs of identical asymmetric centers are present, at least one (for a single pair of identical centers) configuration will be identical with its mirror image; that is, it will possess either a plane or center of symmetry and thus be optically inactive. Such a substance is called a meso compound and is said to be optically inactive because of internal compensation



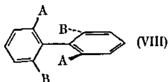
Molecular dissymmetry also derives from structural rigidity, as in the terminally disubstituted allenes (VI) and spiranes (VII), and from the restriction of free rotation about single bonds, as



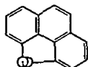
Allene type of asymmetric molecule



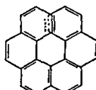
Spiran type of asymmetric molecule



Biphenyl type of asymmetric molecule



4,5-Disubstituted phenanthrene

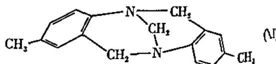


Hexahalicene

Sterically induced types of asymmetric molecule

in the *o,o,o′,o′*-tetrasubstituted biphenyls (VIII). Steric interference with a normally coplanar struc- ture may also produce molecular dissymmetry. For example, phenanthrene with bulky substituents in the 4 and 5 positions (IX) and hexahalicene (X). No asymmetric centers are needed in these types of systems.

When the asymmetric atom is other than carbon, the same general principles apply. In the nitrogen family, optical activity is not observed in tertiary ammonium salts of the type $RR'R''NH^+X^-$, since they readily dissociate to the tertiary amine which undergoes rapid inversions characteristic of ammonia itself. However, in a rigidly bonded system inversion of tertiary nitrogen may be prevented and the nitrogen may then constitute an asymmet- ric center. Thus Tröger's base (XI) has two such asymmetric nitrogens and has been isolated in both optically active configurations.



It is unnecessary to describe in detail the opti- cally active metal-ion complexes, for if the struc- ture of the complex as a whole lacks an *n*-fold al- ternating axis of symmetry, it will be optically active regardless of the nature of the metal ion in the center, which is sometimes said to be asymmet- ric. The complex would remain optically active in many instances (though not in all) even if the metal ion could be removed; only where the posi- tion of the central metal ion produces molecular dissymmetry is its presence essential to optical activity. See ASYMMETRIC SYNTHESIS; COORDINA- TION CHEMISTRY, STEREOCHEMISTRY. [W.R.W.]

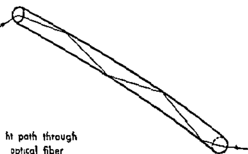
Bibliography. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed., 1957.

Optical fibers

Transparent fibers used to conduct light along se- lected paths. With one fiber provided for each small element of an image, a bundle of fibers will trans- mit a complete image in a manner that is free of many of the restrictions characteristic of conven- tional optical systems

For instance, a fiber bundle may be quite flexible and thus able to convey images around corners or through tortuous channels. This capability makes it well suited to instruments such as the gastro- scope that are designed for the viewing of inacces- sible places. A further unique characteristic is that within the resolving power of the bundle there need be no imaging errors and, by proper arrange- ment of the fibers, rotation, magnification, distor- tion, and field curvature can be completely con- trolled

The image transformations thus made possible are useful, for instance, in stellar spectroscopy where it is desirable to transform a disk-shaped image into a slit-shaped ... her importa


ht path through optical fiber

s of applications of optical fibers utilizes their tive compactness and light-transfer efficiency example, in oscilloscope photography a short, d, fiber bundle used as the tube face can con- t images from the phosphor layer inside the rode ray tube to a photosensitive film in contact h it on the outside. Optical fibers require less ce and operate at higher light-transfer effi- ncy than conventional imaging systems

'or effective light conduction the fibers must be hly transparent with smooth reflective surfaces der these conditions light entering one end is nsmitted to the other end by repeated reflections, shown in the illustration For conduction over tances that are long compared to the fiber diam- r, the large number of reflections makes neces- y a reflection efficiency so high that even well ished metal surfaces are not satisfactory, and phenomenon of total internal reflection is nor- lly used To accomplish this, optical fibers are de of any highly transparent material, usually ss or clear plastic surrounded by another trans- rent material of lower refractive index, which y be simply air

Fiber dimensions are not critical provided the imeters are large compared to the wavelength of ht and provided the fibers in a bundle are spaced least a wavelength apart to prevent light leak- e from one to another Fiber diameters of the der of 0 0005 in are quite feasible thus making ssible the transmission of a standard 525-line levision image through a bundle only about ¼ in ck See REFLECTION (ELECTROMAGNETIC RADIA- ON) [S M M]

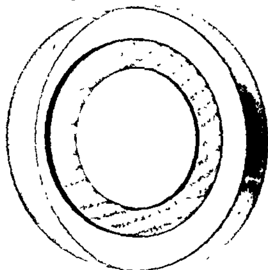
Bibliography John Strong, *Concepts of Classi l Optics* 1958

ptical flat

disk of high grade quartz glass approximately in thick having at least one side ground and olished with a deviation in flatness usually not xceeding 0 000002 in all over, and a surface qual- y of 5 microfinish or less When two surfaces of his quality are placed lightly together so that the ir is not wrung out from between them they will e separated by a film of air and actually touch at nly one point This point will then be the vertex of wedge of air separating the two pieces

If parallel beams of light pass through the flat, part will be reflected against the surface being in spected, while part will be reflected directly back through the flat. Because the distance between the surfaces is constantly increasing along the angle, the beams reflected from the flat and the beams re- flected from the workpiece will alternately rein- force and interfere with each other, producing a pattern of alternate light and dark bands (Fig 1).

matic, the wavelength is known. [illegible] length of 0 0000116 in. is commonly used. Thus a definite relationship is established between lineal measurement and light waves Optical flats are used for two general purposes.


Fig 1 Optical flat being used to determine flatness of seal ring Interference bands on seal ring face show lines of constant depth (The Van Keuren Co)

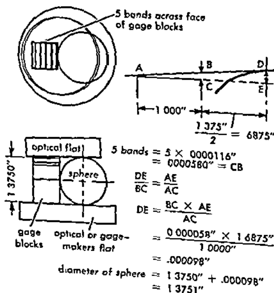


Fig 2 Measurement of height of sphere by means of gage blocks and optical flat

Determination of surface contour. If the surface being inspected is flat, the light bands are parallel. Any deviation from flatness shows as curvature of the lines. The principle in interpreting a pattern is almost identical with the principle in interpreting a topographical map in that the bands connect points of equal distance from the master surface of the optical flat. Deviation from flatness can be reduced to rational figures.

Comparison of lineal measurement. When an optical flat is placed across a gage block or a build-up of blocks and another object, as an end standard, a cylinder or a sphere, both of which are resting on a precisely flat surface, then the angle between the blocks and the flat can be measured and consequently the difference in length between the gage blocks and the length or height or diameter of the unknown can be determined (Fig. 2). With the high degree of accuracy available in electrical and pneumatic comparators, optical flats are seldom used in this way except for spheres or irregular surfaces where point contact by the comparator is impractical.

The principle of interferometry is the standard method for measurement of gage blocks However, an interferometer and not the optical flat itself is used. The wavelength of light is the present standard of all lineal measurements. *See* GAGES; INTERFEROMETRY. [R.A.BO.]

Bibliography: F. H. Rolt, *Gauges and Fine Measurements*, 2 vols., 1929.

Optical materials

The materials used for optical purposes are chiefly glasses, but to a small extent crystalline substances, usually grown for the purpose, are used. In the selection of materials for most uses, the property of being transparent to light is paramount. Glasses transparent to x-rays and to infrared and ultraviolet radiation as well as those opaque to these wavelengths are articles of commerce. Many crystalline materials are transparent to wavelengths to which glass is opaque, as are some plastic materials such as methacrylate (Lucite and Plexiglas) and polystyrene.

Refractive index and dispersion. When radiant energy reaches a boundary between one substance or medium and another, some of it is reflected and the rest passes into the second substance where [illegible] The veloc- [illegible] ly different [illegible] other isotropic materials, the ratio of the velocity in vacuum (v_0) to that in the given substance (v) is called the refractive index of that substance: $n = v_0/v$. *See* REFRACTION OF WAVES.

The absorption of energy in the medium may be substantially uniformly distributed throughout the visible spectrum, in which case the substance appears colorless or gray; or there may be a more or less pronounced maximum of absorption in the visible region, resulting in colored glass The change in the velocity with frequency of the radiant energy may be represented by a dispersion curve in which the refractive index n is plotted against the wavelength λ.

In principle, the shape of the dispersion curve is determined by the quantitative form of the absorption curves in the ultraviolet and the infrared In no case, however, not even for silica glass, has the dispersion curve been completely correlated with known absorptions, and for all practical purposes the various dispersion formulas may be regarded as empirical equations, the constants of which are to be evaluated from measurements of refractive index. *See* ABSORPTION (ELECTROMAGNETIC RADIATION).

The refractive index is usually designated by the letter n, followed by a subscript indicating the wavelength of the light in vacuum; n_D indicates the refractive index for the mean D line of sodium, $\lambda = 589.3$ mμ. In optical glass catalogs, it is customary to give the refractive indices for a group of spectral lines chosen by E. Abbe for the convenience with which they could be obtained for spectrometric work. The source, designation, and wavelength in vacuum of these lines are given in Table 1, together with some other lines which are being used to an increasing extent because of their present greater convenience.

Optical glass catalogs also give the mean dispersion, commonly designated by (C − F) and other partial dispersions, as well as the dispersion ratios (D − C)/(F − C), and so forth, in which the symbols C, D, and F refer to the refractive indices for the spectral lines as designated in Table 1 Another number given under various names is the nu value, $\nu = (n_D - 1)/(n_F - n_C)$.

Optical glass. The optical material most widely used is optical glass. Optical glass, available in a wide range of refractive indices and dispersions, differs from ordinary glass in its freedom from imperfections It must be free from unmelted particles or "stones," from bubbles, and from chemical inhomogeneity, which gives rise to regions of variable refractivity known as cords or striae. Chemi-

Table 1. Designation, source, and wavelength of spectral lines used in spectrometric measurements*

Source	Hg	Hg	H	H	He	Hg
Designation	h	g	G′	F		e
Wavelength, mμ	404.7	435.8	434.1	486.1	492.2	546.1
Source	Na (mean)	He	H	He	K (mean)	
Designation	D	d	C	b	A′	
Wavelength, mμ	589.3	587.6	656.3	706.5	768.2	

* From G. W. Morey, *Properties of Glass*, 2d ed., Reinhold, 1954.

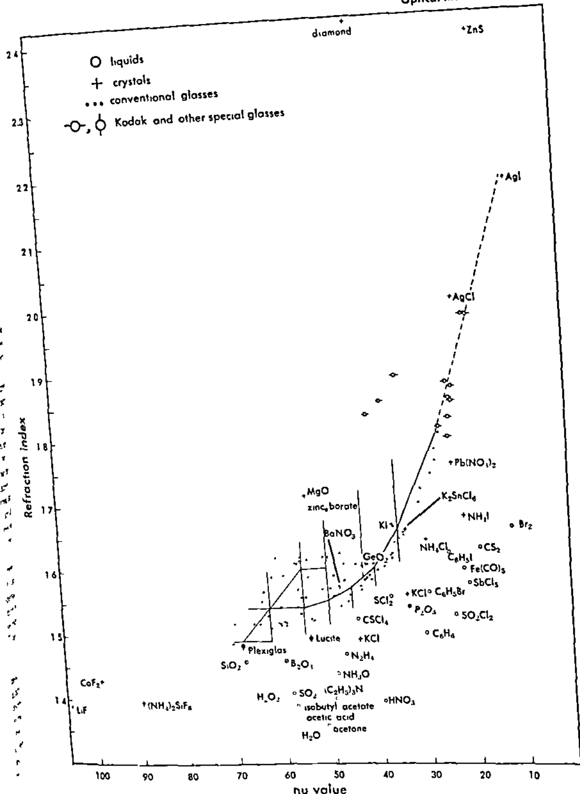


Fig 1 Refractive index and nu values of some liquids, crystals, conventional glasses, and Kodak and other special glasses. Glasses are grouped into types below the line (right to left) are heavy flints, flints, light flints, extra-light flints, short crowns, crowns, borosilicate crowns (triangular area), and fluor crowns, above the line (left to right), phosphate crowns, heavy phosphate crowns, and barium crowns (triangular area), above them heavy barium crowns and light barium flints; above them extra-heavy barium crowns and similar glasses, barium flints, and heavy barium flints, broad field above and to right of heavy barium flints is occupied by Kodak and similar glasses (From G W Morey, *The Properties of Glass*, 2d ed, Reinhold, 1954)

cal homogeneity is obtained by stirring the molten glass, a process discovered about 1790 by a Swiss watchmaker, P.-L. Guinand.

Glass types. The early optical glasses were crowns and flints. The crown glasses were essentially of the same type as window glass, with low index and dispersion, and the flint glasses contained lead oxide and had higher index and dispersion. These glasses, roughly indicated by the full line of Fig. 1, did not have sufficient range in optical properties to enable desirable corrections to be made in optical systems. A great step forward was made by O. Schott in the introduction of the barium crown and flint types, and a second advance was made by G. W. Morey in the rare-earth glasses, also indicated in Fig 1. The range of available glass types and their optical properties are shown in Table 2, which includes wavelengths up to 2.5 μ in the infrared.

Variation of n with temperature. The temperature coefficient of the refractive index is small, ranging from -3×10^{-6} for a fluor crown glass and 9×10^{-6} for silica glass up to 14×10^{-6} for heavy flint glass. The refractive index of silica glass, for example, at 501.6 mμ (He green line), increases from 1.4617 at $-160°C$ to 1.4772 at 1000°C.

Effect of absorption. When the absorption of light in glass is fairly uniformly distributed throughout the visible spectrum, and the amount of absorption is small, the glass appears colorless and limpid as viewed in white light; when the amount of uniform absorption increases, the glass takes on a grayish hue If the absorption is significantly greater for light of any particular color, the transmitted light will appear of a complimentary color; since the absorption is never monochromatic, the color will depend on the thickness of the sample. Comparisons are best made by means of curves showing the change of absorption or transmission with wavelength for samples of standard thickness.

Silica glass, in thicknesses such as are used in cameras and optical instruments, transmits all the radiation to which ordinary photographic plates are sensitive, that is, down to a wavelength of 220 mμ, but wavelengths of 193 mμ and less are almost completely absorbed The transmission of the usual colorless glass is limited chiefly by absorption bands, as is indicated by the dispersion curve. The limit of transmission in ultraviolet light is determined largely by the content of Fe_2O_3, which shows strong absorption in the near ultraviolet The limit of transmission in the very near infrared is determined largely by the content of FeO, which shows strong absorption at about 1 μ The best optical glasses have a transparency of over 99% throughout the visible spectrum (380–780 mμ), a result achieved only by using the greatest care in excluding impurities, especially iron. Ordinary window glass removes the 313-mμ line of mercury vapor and shorter wavelengths. Typical transmission curves of a number of glasses are shown in Fig 2

Colored glass. The coloring agent or colorant in glass usually is considered to be produced by (1) substances dissolved in the glass which absorb characteristic frequencies, (2) particles of submicroscopic dimensions, colloidally dispersed in the glass, such as gold or copper in ruby glass, or

Table 2. Refractive indices of some optical glasses*

Glass no	1	2	3	4	5	6	7	8	9	10	11
Maker	B-L†	B-L	PM‡	B-L	B-L	PM	B-L	PM	B-L	Schott	B L
Type no	70	20–2	5951	160–1	110–1	6581	120–1	4887	120–2	F2	220-3
Name	Hard crown	Borosilicate crown	Extra-light flint	Telescope flint	Dense barium crown	Light flint	Dense barium crown	Light barium flint	Dense barium crown	Dense flint	Extra-dense flint
n_D	1.51166	1.51666	1.52390	1.52970	1.57136	1.57348	1.61173	1.60535	1.61703	1.62082	1.71968
$n_F - n_C$	0.00845	00802	01039	01028	00995	01338	01037	01381	01141	01707	02451
ν	60.6	61.4	50.4	51.5	57.4	42.9	59.0	43.8	54.1	36.4	29.4
Wavelength											
0.3650 μ	1.53245	(?)	1.55072	1.55555	1.59610	1.60880	1.63738	1.64171	1.64570	1.66714	(absorbed)
h 4046	1.52597	1.53013	1.54192	1.54736	1.58837	1.59718	1.62935	1.62975	1.63666	1.65163	1.76506
G′ 4340	1.52239	1.52676	1.53728	1.54286	1.58407	1.59098	1.62491	1.62338	1.63164	1.64316	1.75270
g 4358	1.52217	1.52657	1.53703	1.54262	1.58382	1.59064	1.62467	1.62300	1.63138	1.64300	1.75205
F 4861	1.51760	1.52227	1.53125	1.53695	1.57838	1.58300	1.61903	1.61516	1.62508	1.63307	1.73732
e 5461	1.51375	1.51864	1.52645	1.53224	1.57380	1.57678	1.61429	1.60870	1.61982	1.62503	1.72569
D 5893	1.51166	1.51666	1.52390	1.52970	1.57136	1.57348	1.61173	1.60535	1.61703	1.62082	1.71968
C 6563	1.50915	1.51425	1.52086	1.52667	1.56843	1.56962	1.60866	1.60135	1.61367	1.61600	1.71278
60	1.51123	1.51623	1.52336	1.52913	1.57081	1.57280	1.61118	1.60456	1.61638	1.62005	1.71842
65	1.50937	1.51447	1.52112	1.52688	1.56867	1.56997	1.60891	1.60163	1.61394	1.61632	1.71330
70	1.50789	1.51302	1.51934	1.52507	1.56690	1.56764	1.60709	1.59930	1.61196	1.61338	1.70935
75	1.50659	1.51179	1.51788	1.52357	1.56542	1.56583	1.60557	1.59740	1.61030	1.61126	1.70608
80	1.50548	1.51070	1.51658	1.52230	1.56420	1.56422	1.60423	1.59581	1.60890	1.60936	1.70358
85	1.50458	1.50978	1.51548	1.52117	1.56314	1.56290	1.60312	1.59445	1.60773	1.60780	1.70140
90	1.50373	1.50893	1.51451	1.52015	1.56217	1.56172	1.60212	1.59324	1.60668	1.60635	1.69956
95	1.50299	1.50818	1.51367	1.51927	1.56137	1.56066	1.60123	1.59222	1.60576	1.60517	1.69794
1.00	1.50229	1.50745	1.51289	1.51844	1.56063	1.55975	1.60039	1.59128	1.60493	1.60408	1.69651
1.5	1.49665	1.50114	1.50677	1.51158	1.55465	1.55282	1.59380	1.58464	1.59834	1.59632	1.68725
2.0	1.49094	1.49446	1.50096	1.50460	1.54896	1.54647	1.58732	1.57887	1.59268	1.59026	1.68010
2.5	1.48394	1.48597	1.49393	1.49588	1.54197	1.53901	1.57931	1.57218	1.58565	1.58114	1.67315

* From G W Morey, *Properties of Glass* 2d ed, Reinhold 1954 prepared from data by R Kingslake and H G Conrady

† Bausch and Lomb

‡ Parra-Mantois

Fig 2 Transmission curves of glasses for controlling the ultraviolet portion of the spectrum, made by Corning Glass Works, and identified by Corning's color specification number. Number 0–53 is standard chemical Pyrex brand glass, 0–54 shows a blue fluorescence when excited by wavelength of 254 mμ, 7–59 shows 15% transmittance of 499 mμ (Corning Glass Works)

(3) particles of microscopic or larger dimensions, either themselves colored, as in the aventurine glasses, or colorless, as in opal glass The coloring agents which act by virtue of characteristic absorption spectra are all elements belonging to the transition rows of the periodic system and especially to the first of these rows to which belong titanium vanadium chromium, manganese iron, cobalt, nickel and copper, the commonest and most effective colorants of glass

Crystalline materials. A number of crystalline materials are available for use as optical materials The mineral Iceland spar, a well-crystallized $CaCO_3$, is used as a polarizing material because of its high birefringence (double refraction) It is a uniaxial mineral, and is used only for special purposes See BIREFRINGENCE; CRYSTAL OPTICS

A number of manufactured isotropic crystalline substances have become available These are noteworthy for their high transmissions in the ultraviolet and infrared

Sodium chloride, NaCl, is the same as the mineral halite or rock salt It is used in ultraviolet, visible, and infrared spectroscopy, in infrared microspectrography, and for lens elements in microscope objectives for use in the infrared or in the ultraviolet Potassium chloride, KCl (sylvite), is also used for windows and prisms in ultraviolet and infrared spectroscopy.

Other crystals especially useful at high or low frequencies are CaF_2, MgO, LiF, KBr, CsBr, CsI, AgCl, and a solid solution of 50 mole % TlBr and 50 mole % TlI, usually called KRS-5. Silver chloride (AgCl) is insoluble in water and may be used for windows of cells containing aqueous solutions. See OPTICAL FIBERS. [C.W.MO.]

Bibliography: G. W Morey, *The Properties of Glass*, 2d ed., 1954; R. B Sosman, *The Properties of Silica*, 1927; W. A. Weyl, *Coloured Glasses*, 1951

Optical methods of chemical analysis

These methods deal with the measurement of the extent of the interaction of light (electromagnetic radiation) with matter. Both the manner of interaction and the type of electromagnetic radiation utilized vary widely. The common wavelength units are the angstrom (A), millimicron (mμ), and micron (μ).

$$1\ A = 10^{-8}\ cm$$
$$1\ m\mu = 10^{-7}\ cm$$
$$1\ \mu = 10^{-4}\ cm$$
Thus $$1\ \mu = 1000\ m\mu = 10{,}000\ A$$

Frequency units are usually given in cm^{-1}, called wave numbers or kaysers, so that the frequency corresponding to 1 μ is 10,000 cm^{-1}.

A broader interpretation of optical methods also includes the corresponding techniques using higher energy x rays and lower energy microwaves. For discussions of the principal methods see COLORIMETRIC ANALYSIS, FLAME PHOTOMETRY; FL[illegible] ELECTROCHEMICAL ANALYSIS; SPECTROPHOTOMETRIC ANALYSIS, TURBIDIMETRIC ANALYSIS; X-RAY FLUORESCENCE ANALYSIS [R F G.]

Bibliography: H. H Willard, L. L. Merritt Jr., and J A Dean, *Instrumental Methods of Analysis*, 3d ed., 1958

Optical recording

The process of recording sound signals on photographic film so that they may be reproduced at a subsequent time. Optical recording is also termed motion-picture recording or photographic recording The narrow bands on motion-picture film used for the sound record are called sound tracks. This article discusses the important systems, including stereophonic systems, used in the recording and reproducing of sound in the motion-picture industry

Monophonic system. A monophonic sound [illegible] tion-picture [illegible]
a modulator
and a mecha

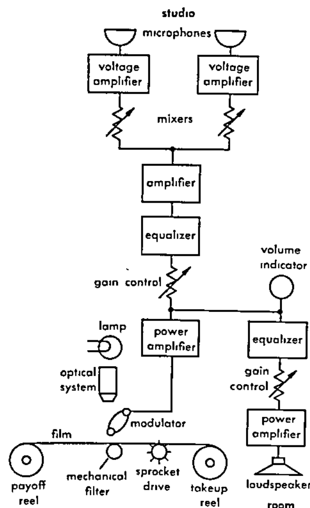


Fig. 1. Schematic arrangement of apparatus in a complete optical sound motion-picture recording system.

graphic film relative to the light beam and thereby recording signals on the film corresponding to the electrical signals A monophonic sound motion-picture reproducing system is basically a combination of a light source, optical system, photoelectric cell, and a mechanism for moving a film carrying an optical record by means of which the recorded photographic variations are converted into electrical signals of approximately like form

Recording system. The elements in a complete monophonic sound motion-picture recording system are shown in Fig 1. The output of each microphone is amplified and fed to a mixer, a device having two or more inputs and a common output If more than one microphone is used, as for example when there are two actors, one microphone for each, the outputs of the two microphones may be adjusted for the proper balance by means of the mixers. An electronic compressor is used to reduce the amplitude range to that suitable for reproduction in the home. An equalizer provides the standard motion-picture recording characteristic (*see* EQUALIZATION, FREQUENCY-RESPONSE). The gain control provides means for controlling the over-all signal level fed to the power amplifier. The light modulator, actuated by the amplifier, records a photographic image upon the film corresponding to the electrical input. A monitoring system consisting of a volume indicator, a complementary equalizer, gain control, power amplifier, and loudspeaker or headphone is used to control the recording operation.

Modulator. In the variable-area recording system the transmitted light amplitude is a function of the amount of unexposed area in the positive print This type of sound track is produced by means of a mirror galvanometer which varies the width of the light slit under which the film passes. The elements of a variable-area recording system are shown in Fig. 2. The triangular aperture is uniformly illuminated by means of a lamp and lens system. The image of this triangular aperture is reflected by the galvanometer mirror focused on the mechanical slit, which in turn is focused on the film The galvanometer mirror swings about an axis parallel to the plane of the paper. The triangular light image on the mechanical slit moves up and down on the mechanical slit. The result is that the width of the exposed portion of the negative sound track corresponds to the rotational vibrations of the galvanometer. In the positive record, the width of the unexposed portion corresponds to the signal.

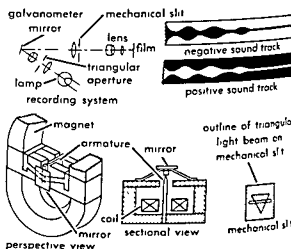


Fig 2 Elements of a variable-area sound motion-picture film recording system.

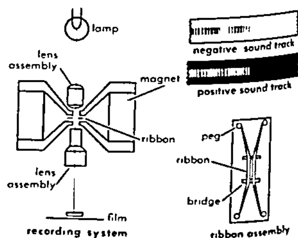


Fig 3 Elements of a variabl … lty sound motion-picture film recording syste

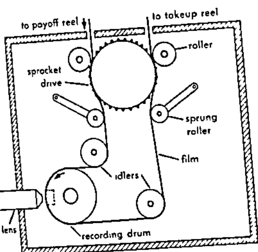


Fig 4 Schematic view of the photographic film-transport mechanism in a motion-picture film sound recorder (After H F Olson, Acoustical Engineering, 3d ed, Van Nostrand, 1957)

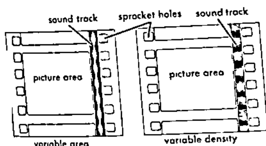


Fig 5 Position of the picture and sound track in a 35-mm sound motion-picture film Two types of sound track, variable-area and variable-density, are shown (After H F Olson, Acoustical Engineering, 3d ed, Van Nostrand, 1957)

In the variable-density system the transmitted light amplitude is an inverse function of the amount of exposure in the positive print. This type of sound track is produced by means of a light valve which varies the amount of light falling upon the moving film The elements of a variable-density recording system are shown in Fig 3. The ribbons of the light valve are illuminated by means of a lamp and lens system The image of the illuminated slit produced by the ribbons of the light valve is focused on the film The amount of exposure on the negative film varies with the aperture at the ribbons, whereas in the positive record the amount of exposure is an inverse function of the input to the light valve.

Recording film transport The film-transport mechanism used in recording sound on film consists of a positive drive of the perforated film and a constant speed drive of the film where the modulated light beam strikes the film. A film transport mechanism of this type is shown in Fig 4. Positive drive of the film is obtained by the sprocket drive, which is interlocked with the camera drive so that synchronism of picture and sound will be obtained When the film passes over the sprocket drive, variations in the motion of the film at the sprocket-hole frequency are produced. These variations in the film speed must be removed at the recording point to eliminate spurious frequency modulation of the image on the film. Uniform speed at the recording point is provided by a mechanical

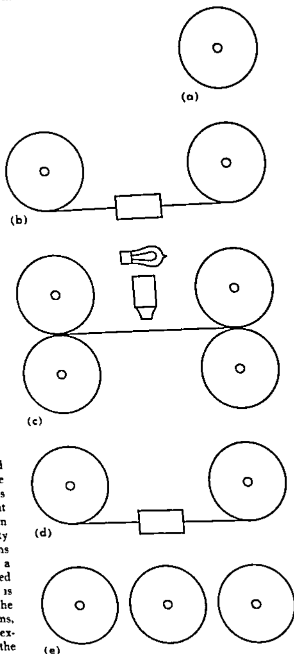


Fig 6 The steps in the process for the production of motion-picture positive film from the negative film. (a) Undeveloped negative record on film (b) Negative developer (c) Developed negative record on film printer Printed positive record on film (d) Positive developer (e) Positive film records on reels (After H F. Olson, Acoustical Engineering, 3d ed., Van Nostrand, 1957)

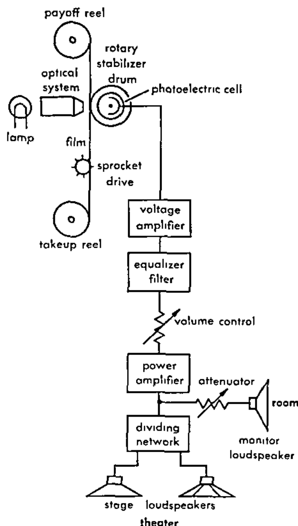


Fig 7 Schematic arrangement of apparatus in a complete optical sound motion-picture reproducing system.

contrivance called a filter. It is located between the sprocket drive and recording point and consists of the inertia of the recording drum and the compliance of the film between the recording drum and the sprocket drive The recording drum is driven by a magnetic system from the motor which drives the sprocket and thereby imparts a slight amount of drive to the film The magnetic drive isolates the variations in the rotational speed of the motor drive from the rotating drum. The combination of the isolating filter and magnetic drive provides a system with very uniform motion of the surface of the drum. The image of the modulator is focused on the film while it is in contact with the drum.

Film and sound track. In the recording of sound motion pictures, the picture and sound are recorded on separate photographic films. Therefore, the camera and sound recorder must be synchronized This is accomplished by the use of an interlock system between the camera and sound recorder and the use of perforated film in the form of sprocket holes along the two edges of the film for both the camera and sound recorder.

The sound track on 35-mm film occupies a space about 0.1 in. wide just inside the sprocket holes, as shown in Fig. 5. There are two types of sound track in general use today, variable-area and variable-density. The type of variable-area sound track shown in Fig. 5 is termed bilateral variable-area.

Film developing and printing. The processes used in the film laboratory for the mass production of motion-picture positive prints are shown in Fig. 6. The negative record of Fig. 6*a* is developed as shown in Fig. 6*b*. Then the required number of positive prints of both picture and sound is printed from the negative record as shown in Fig 6*c*. These positive records are developed as shown in Fig. 6*d*. The positive records of Fig. 6*e* are used for sound reproduction and picture projection in the theater.

Reproducing system. The elements in a complete monophonic sound motion-picture reproducing system are shown in Fig. 7. The first element is the optical system consisting of a lamp and a lens arrangement which produces an illuminated slit of light upon the film. The light beam passes through the film and falls upon the photoelectric cell. When the film is pulled past the slit the variations in light, which are due to the variable-density or variable-area recording on the film, fall upon the photoelectric cell and are converted into the corresponding electrical variations. The output of the photoelectric cell is fed to an amplifier followed by a filter, which is used to cut the ground noise (residual noise in the absence of the signal) due to the film above the upper limit of reproduction, and by equalizers, which are used to adjust the frequency characteristic to that suitable for the best sound reproduction in the theater. The volume control (gain control) is used for adjusting the level of sound output. The output of the power amplifier feeds the stage loudspeakers, located behind the screen, and the monitoring loudspeaker. Except for the stage loudspeakers, the entire equipment including the monitoring loudspeaker is located in the projection booth.

Optical electronic reproducer. The elements of a motion-picture film sound reproducing system are shown in Fig 8 The light source, in the form of an incandescent lamp, is focused upon a mechanical slit by means of a condensing lens. The mechanical slit in turn is focused on the negative film The height of the image on the film is usually about 0.00075 in Under these conditions the amount of

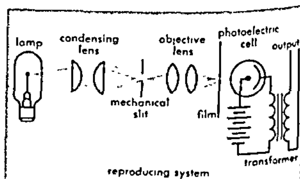


Fig 8 Elements of a motion-picture film sound reproducing system

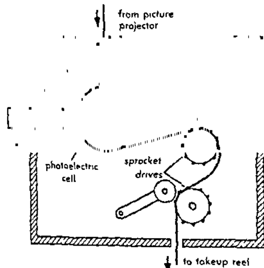


Fig 9 Schematic view of photographic film transport mechanism of a motion-picture film sound reproducer (After H F Olson, *Acoustical Engineering*, 3d ed., Van Nostrand, 1957)

light which impinges upon the photocell is proportional to the unexposed portion of the sound track in variable-area recording or to the inverse function of the density in variable-density recording When the film is in motion, the resultant light undulations which fall upon the photocell correspond to the voltage variations applied to the recording galvanometer The voltage output of the photocell is proportional to the amount of light which falls upon the cathode

Reproducing film transport The film transport used in reproducing sound on photographic film consists of a positive drive of the perforated film and a constant speed drive where the light passes through the film to the photoelectric cell A film-transport mechanism of this type is shown in Fig 9 Positive drive of the film is obtained by means of the two sprocket drives The sprocket drives are geared with the positive picture drive so that a constant loop of film is maintained between

... loop of film between the picture head and sound head, and for this reason variations in the picture drive will not be imparted to the sound head

After the film enters the sound head it passes over a drum. The light beam of the reproducing system passes through the film to the photocell located inside the drum while the film is on the drum The drum is driven by the first sprocket drive. The compliance of the film between the film and the sprocket provides a mechanical filter system and thereby reduces the sprocket-hole ripple at the drum Under these conditions, the drum is rotated at a constant speed, and as a consequence the film will move past the light beam at a constant speed. The second sprocket isolates the takeup reel from the reproducing system

Distortion and noise in reproduction. Most commonly, distortion in an optical reproducing system is due to the inherent nonlinear characteristics of the photographic process. This type of distortion can be reduced to a low value by the use of proper illumination in the recording or duplicating process The developing processes must also be accurately controlled in order to achieve a low value of nonlinear distortion.

Noise in clean film is due to the inherent grain structure of the photographic medium. Scratches and foreign particles on the film add additional noise

Another ...

tic ...

pr ...

tion of the reproduced signal and is termed flutter and wow See FLUTTER AND WOW

Stereophonic system. Stereophonic sound reproduction employing multiple channels has been introduced on a wide scale in connection with wide-screen motion pictures In one system, three separate magnetic-tape channels are used for the reproduction of stereophonic sound. The three-channel stereophonic sound system for recording the sound on magnet...

...trols of the three channels are ganged so that the same amplification is maintained in all channels, and an equalizer provides the desired recording

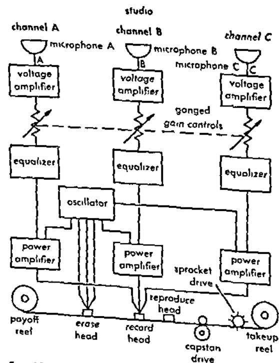


Fig 10 Schematic arrangement of apparatus in a complete three-channel stereophonic magnetic-tape sound motion-picture recording system

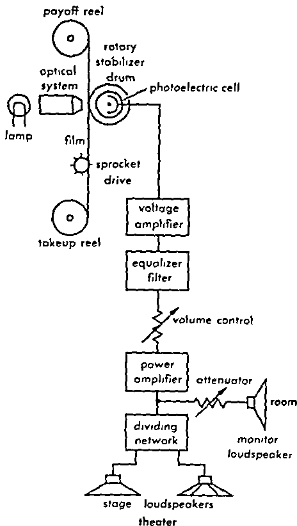


Fig. 7. Schematic arrangement of apparatus in a complete optical sound motion-picture reproducing system.

contrivance called a filter. It is located between the sprocket drive and recording point and consists of the inertia of the recording drum and the compliance of the film between the recording drum and the sprocket drive The recording drum is driven by a magnetic system from the motor which drives the sprocket and thereby imparts a slight amount of drive to the film The magnetic drive isolates the variations in the rotational speed of the motor drive from the rotating drum The combination of the isolating filter and magnetic drive provides a system with very uniform motion of the surface of the drum. The image of the modulator is focused on the film while it is in contact with the drum

Film and sound track. In the recording of sound motion pictures, the picture and sound are recorded on separate photographic films. Therefore, the camera and sound recorder must be synchronized. This is accomplished by the use of an interlock system between the camera and sound recorder and the use of perforated film in the form of sprocket holes along the two edges of the film for both the camera and sound recorder.

The sound track on 35-mm film occupies a space about 0.1 in. wide just inside the sprocket holes, as shown in Fig. 5. There are two types of sound track in general use today, variable-area and variable-density. The type of variable-area sound track shown in Fig. 5 is termed bilateral variable area.

Film developing and printing. The processes used in the film laboratory for the mass production of motion-picture positive prints are shown in Fig. 6. The negative record of Fig. 6a is developed as shown in Fig. 6b. Then the required number of positive prints of both picture and sound is printed from the negative record as shown in Fig. 6c. These positive records are developed as shown in Fig. 6d. The positive records of Fig. 6e are used for sound reproduction and picture projection in the theater.

Reproducing system. The elements in a complete monophonic sound motion-picture reproducing system are shown in Fig. 7. The first element is the optical system consisting of a lamp and a lens arrangement which produces an illuminated slit of light upon the film. The light beam passes through the film and falls upon the photoelectric cell. When the film is pulled past the slit the variations in light, which are due to the variable-density or variable-area recording on the film, fall upon the photoelectric cell and are converted into the corresponding electrical variations. The output of the photoelectric cell is fed to an amplifier followed by a filter, which is used to cut the ground noise (residual noise in the absence of the signal) due to the film above the upper limit of reproduction, and by equalizers, which are used to adjust the frequency characteristic to that suitable for the best sound reproduction in the theater. The volume control (gain control) is used for adjusting the level of sound output The output of the power amplifier feeds the stage loudspeakers, located behind the screen, and the monitoring loudspeaker. Except for the stage loudspeakers, the entire equipment including the monitoring loudspeaker is located in the projection booth

Optical electronic reproducer. The elements of a motion-picture film sound reproducing system are shown in Fig. 8. The light source, in the form of an incandescent lamp, is focused upon a mechanical slit by means of a condensing lens. The mechanical slit in turn is focused on the negative film. The height of the image on the film is usually about 0.00075 in. Under these conditions the amount of

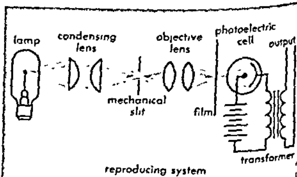


Fig 8. Elements of a motion-picture film sound reproducing system.

The instruments used fall by function into two classes, those which determine spatial position and those which record engineering events. Tracking telescopes are the basic engineering-event recording systems, while cinetheodolites and ballistic cameras are used for the precise determination of spatial position

Spatial position determination. This can be considered as the triangulation of a moving target using the methods of the civil surveyor. The target

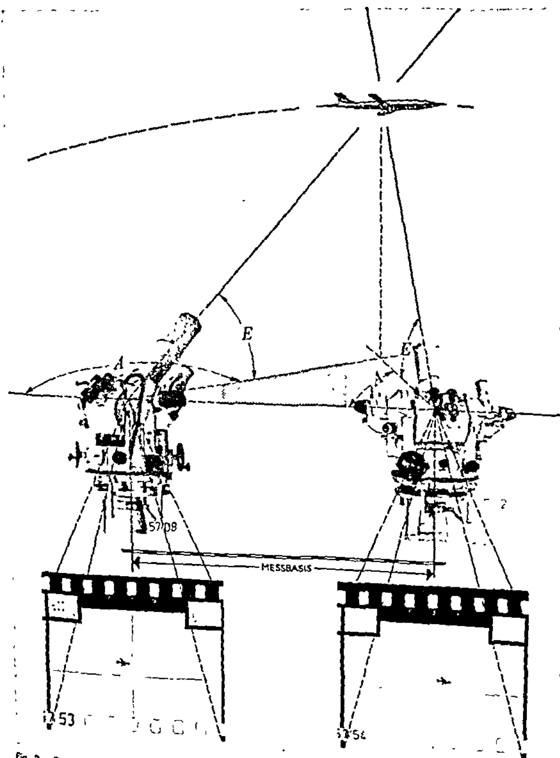


Fig 2 Cinetheodolite triangulation The cinetheodolites are on a precisely surveyed base line The missile shown in the typical frame is not precisely on the optical axis Angles E and A must be corrected for this offset before triangulating for position of missile

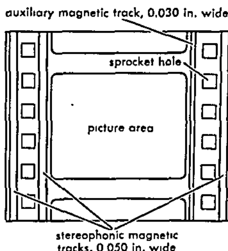


Fig. 11. Magnetic strips on a motion-picture positive film.

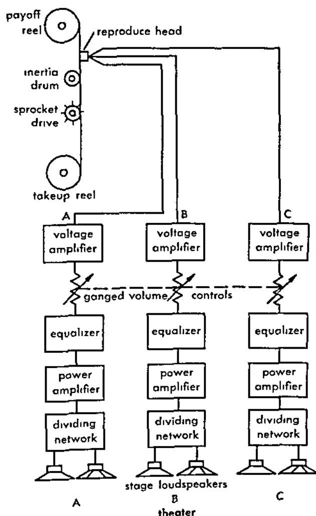


Fig 12. Schematic arrangement of apparatus in a complete three-channel stereophonic magnetic-tape sound motion-picture reproducing system.

characteristic. The output of the power amplifier is fed to the recording head. Positive drive of the film is obtained by the sprocket drive, which is interlocked with the camera drive so that synchronism of the picture and sound will be obtained

In the release film prints, the information is recorded on magnetic-tape strips cemented to the positive film (Fig. 11). Note that there are four strips. In some instances the fourth channel representing the information on the fourth strip is used for control or auxiliary information.

The elements of a three-channel magnetic-tape stereophonic sound motion-picture reproducing system are shown in Fig. 12. The output of the three-channel magnetic head is fed to three separate voltage amplifiers. The volume controls of the three channels are ganged so that the same amplification is maintained in the three channels. The equalizers are used to adjust the frequency characteristics to those suitable for the best reproduction in the theater. The output of the power amplifiers feeds the three sets of stage loudspeakers, which are located behind the screen. *See* CINEMATOGRAPHY, MAGNETIC RECORDING. [H F O]

Bibliography: Academy of Motion Picture Arts and Sciences Research Council, *Motion Picture Sound Engineering*, 1938; J. G. Frayne and H. Wolff, *Elements of Sound Recording*, 1949, H F Olson, *Acoustical Engineering*, 3d ed , 1957.

Optical tracking instruments

A family of optical instruments used for precise time-correlated observation of distant airplanes, missiles, and artificial satellites, all of which travel at apparent velocities much greater than those of most astronomical objects. The instruments supply permanent engineering records for the determination of spatial position, missile attitude, structural behavior, and performance of specific mechanisms during test flights. These observations enable engineers to correct design, improve performance, and collect scientific data from missiles at extreme distances and altitudes.

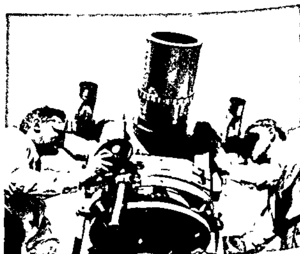


Fig. 1. Bodenseewerk cinetheodolite with 100-cm objective Operators are tracking the missile with 10-power sighting telescope Geared handwheels are used to control instrument motion. Electronically aided tracking drives have been developed which allow more precise control of the cinetheodolite motion (Bodenseewerk Perkin-Elmer and Co)

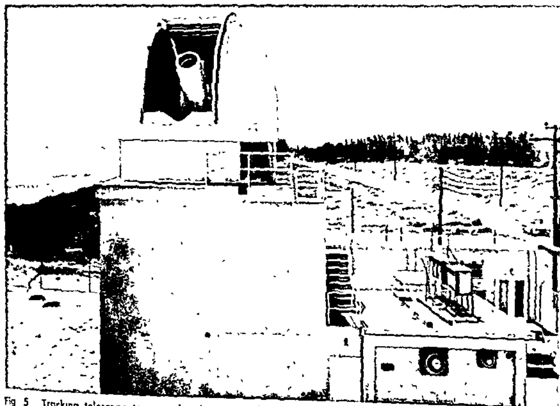


Fig 5 Tracking telescope tower with telescope and dome in operating position The interior of the tower houses the electronic control room, photographic darkroom, and maintenance work area. (Perkin-Elmer Corporation)

ical stability and accuracy, allowing ballistic cameras to achieve 2–5 sec angular accuracy

Pyrotechnic flares or electronic stroboscopic lamps at the missile are used to indicate the missile positions against the night sky The image of each flare is measured with respect to the surrounding stars The use of lenses of less than 10-micron distortion allows the lines of sight from camera to missile to be determined with 25 reference stars The lines of sight are used to determine the missile positions by methods similar to those used with cinetheodolites.

The high accuracy of ballistic cameras is the result of the static mode of operation The camera shutter remains open for the entire time of passage of the missile across the field of view The photographing on a single plate of more than 100 missile points against the star field permits precise position, velocity, and acceleration measurements without degradation by the flexure, mislevel, and vibration of tracking motion. The location of both the stars and missile outside the earth's atmosphere decreases the effect of geometric distortion caused by the uncertainty of atmospheric refraction corrections.

The disadvantages of ballistic cameras are the requirement for night operation, the dependence on missile borne flares or lamps, and the comparative difficulty of reducing the data.

The ballistic camera most commonly used at test ranges is the Wild BC-4 with an 18 × 18-cm field size and lenses from 30 5-mm focal length at $f/2\,6$, to 11 5-cm focal length at $f/5\,6$ The camera and support system are equipped with vertical and horizontal optically graduated circles for the precise orientation of the camera prior to operation. A camera capping shutter is supplied for the time correlation of the stellar calibration exposures.

For daytime operation without stellar calibration, the camera is equipped with a between-the-lens shutter for 6 to 30 exposures of 0 004 sec. To prevent overexposure during the multiple exposures, a focal-plane sky screen is incorporated. It can be controlled by a master tracking instrument. The result is that each exposure illuminates only the small area of the plate which includes the missile The final daytime plate is a series of small exposed areas, each including an image of the missile The daytime mode of operation is sensitive to the same field effects as cinetheodolites, since the camera orientation is established by level vials and optically graduated circles

Ballistic camera planning for the future is centered about the use of larger-aperture systems to increase the operating range of the cameras while simultaneously decreasing the intensity of the electronic or pyrotechnic missile-borne sources. The

motion requires that a minimum of two instruments separated by a base line be used. Each instrument records the data for computing the direction of the line of sight to the target for each instant of time. This information, in the form of elevation and azimuth angles, the times of observation, and the known location of each instrument, is used to triangulate for the location of the missile as a function of time.

Most of the spatial position work is performed by cinetheodolites (Fig. 1), which are surveying theodolites having 35-mm motion-picture cameras with 60- to 200-cm focal-length lenses substituted for the surveyor's eye and telescope. The system of cameras is pulsed up to a maximum of 10 frames/sec from a master control station for simultaneous exposure as the cinetheodolites follow the moving missile. Each 29×36.5-mm photograph records the elevation angle E, the azimuth angle A, the missile image, and the reticle lines which define the instrumental axis (Fig. 2). The angles are measured with optically graduated circles accurate to 2 seconds of arc. The tracking error (the difference between the instrumental line of sight and the line of sight to the missile) is determined from the missile photograph. The tracking error is then used to determine the missile line of sight from the recorded instrumental angles. Photography of known targets and precise leveling are used to establish the relationship between the instrument angles and the test coordinate system.

During the procedure, two operators follow the missile visually through two 10- to 30-power sighting telescopes, guiding the cinetheodolite by turning geared handwheels as required to keep the cinetheodolite axis pointed at the missile.

A minimum of three cinetheodolites is used to ensure optimum triangulation as the missile-instrument geometry changes with missile motion. The accuracy of cinetheodolite systems under field conditions is 15–30 sec of arc. The degradation of the 5-sec laboratory accuracy is the result of thermal and dynamic deformation under field-tracking conditions.

The requirement for cinetheodolites capable of smoother, more precise tracking of missiles at increased altitudes and velocities has led to the development of massive, electrohydraulic, servo-driven cinetheodolites such as the 70-mm Recording Optical Tracking Instrument (ROTI) [illegible] two 16-in.-aperture telescopes, [illegible] focal length, [illegible] focal length for wide-angle coverage. *See* PHOTOGRAPHY, SCHMIDT CAMERA; TELESCOPE, ASTRONOMICAL.

Ballistic cameras. These are fixed-axis, wide-angle, photographic-plate cameras, capable of more precise spatial position determination by recording on one plate multiple exposures of the missile against a stellar background (Fig. 3). The use of a static system and precisely cataloged star positions decreases the necessity for long-term mechan-

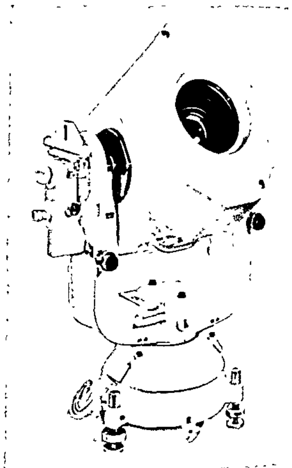


Fig. 3 Ballistic camera, Wild BC-4, with 30.5-mm, f/2.6 lens and 18 × 18-cm picture size. Graduated circles and level vials are used to orient camera prior to test. During night photography of missile, shutter remains open with camera locked at one position (Wild Heerbrugg Instruments, Inc.)

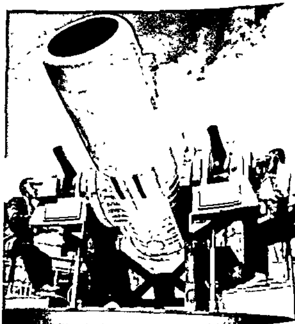


Fig. 4 Tracking telescope, ROTI Mk II. During partially overcast conditions, the telescope can be slaved to follow the parallax-converted data from a radar unit (Perkin-Elmer Corporation)

So great were the contributions of Newton to optics that a hundred years went by before further outstanding discoveries were made. In the early nineteenth century many productive investigators, foremost among them Thomas Young and A. J. Fresnel, established the transverse-wave nature of light. The relationship between optical and magnetic phenomena, discovered by M. Faraday in the 1840s, led to the crowning achievement of classical optics—the electromagnetic theory of J. C. Maxwell (*see* ELECTROMAGNETIC RADIATION; LIGHT; MAXWELL'S EQUATIONS). Maxwell's theory, which holds that light ...

ment of optical phenomena. In particular, it served as the basis for understanding the interaction of light with matter, and hence as the basis for treatment of the phenomena of physical optics. In the hands of H. A. Lorentz, this treatment led, at the end of the last century and the beginning of the present, to an explanation of many optical phenomena, such as the Zeeman effect, in terms of atomic and molecular structure. The theories of Maxwell and Lorentz are regarded as the culmination of classical optics.

In the present century optics has been in the forefront of the revolution in physical thinking caused by the theory of relativity and especially by the quantum theory. To explain the wavelength dependence of heat radiation, the photoelectric effect ...

... mechanism of the interaction of radiation and matter and about the nature of radiation itself has been found necessary. The chief early quantum theorists were M. Planck, A. Einstein, and N. Bohr; later came L. de Broglie, W. Heisenberg, P. A. M. Dirac, E. Schrodinger, and others.

At present the science of optics finds itself in a position that is satisfactory for practical purposes but less so from a theoretical standpoint. The theory of Maxwell is sufficiently valid for treating the interaction of high-intensity radiation with systems considerably larger than those of atomic dimensions. The modern quantum theory is adequate for an understanding of the spectra of atoms and molecules, and for the interpretation of phenomena ...
one do...
the pr...
Howev...
electro...
tems h... yet been worked out.

[R C L.]

Bibliography: M. Born, *Optik*, 1933; M. Born and E. Wolf, *Principles of Optics*, 1959; F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed., 1957; J. A. Stratton, *Electromagnetic Theory*, 1941; J. Strong, *Concepts of Classical Optics*, 1958; R. W. Wood, *Physical Optics*, 3d ed., 1934.

Optics, geometrical

The geometry of light rays and their imagery through optical systems. The phenomena of diffraction due to the finite apertures of the lens systems are neglected in geometrical optics.

Reflection and refraction laws. Light moves in straight lines through homogeneous media and changes its direction at the ...

one is reflected back into the first medium and the other penetrates the second medium after being bent or refracted. The incident, reflected, and refracted rays all lie in one plane containing the surface normal and form angles i, i_r, and i' with the surface normal such that

$$i_r = \pi - i \quad (1)$$

and

$$n \sin i = n' \sin i' \quad (2)$$

where n and n' are the refractive indices of the media separated by the refracting surface. In order to obtain a solution of Eq. (2), i and i' must be chosen so that they are in the same quadrant. *See* REFLECTION (ELECTROMAGNETIC RADIATION); REFRACTION OF WAVES; *see also* LIGHT.

These formulas together with pure geometry make it possible to trace a ray through a system of lenses. The specific form of the ray tracing formula should be adapted to the tools of the lens designer, that is, it will be different for a person using logarithm tables and for one using an electric desk machine or an electronic computer.

Point source. A point source is either an artificial light source which is so small that it appears to a given optical system as a point, or a luminous object, such as a star, which is so far away that it sends out coherent light.

All physical objects have finite areas. However, because of diffraction at the aperture of an optical system, or of the eye, an object which is small compared with the Airy disk will be imaged as an Airy disk, that is, it will be indistinguishable from the theoretical image of a mathematical point. Such an object can be given as the definition of a physical point. For a discussion of the Airy disk see DIFFRACTION.

Characteristic function. The aim of the optical designer is to see what happens to all the rays coming from every point of the object to be imaged. Moreover, he wants to direct the rays so that all of them coming from a fixed object point are collected at a fixed image point (freedom from aberration). Mostly, he wants all these image points to lie on a plane (freedom from field curvature), and he wants the image to be similar in shape to the object (freedom from distortion). Finally, correction should be achieved for light of ...

... investigating all these problems is the characteristic function. If a coordinate



Fig. 6. Thor missile photographed at 35 miles using 24-in.-diameter, 500-in.-focal-length tracking telescope. (U S. Air Force)

use of larger-aperture cameras will also decrease the atmospheric refraction errors which presently limit ballistic camera accuracy to 2 sec of arc

The largest refracting lens being considered is of 600-mm focal length, f/2.0 relative aperture. This lens is considered the largest refracting system feasible for use in a ballistic camera It is expected that for apertures greater than this there will be increasing use of reflecting systems similar to the f/0 67, 8-in. focal length, Baker-Super-Schmidt meteor camera and the 20-in f/1.0 satellite tracking camera. A system of 40-in. focal length at f/0.9 is being considered The systems are distortionless but do not have a flat focal plane. Projectors can be constructed to project accurately the curved plates onto flat glass prints for convenient, precise measurement

Tracking telescopes. These are long-focal-length telescopes mounted to track missiles in flight precisely while collecting missile performance data The first systems were crude attempts to track manually with 35-mm cameras of 12- to 24-inch focal length. Increased focal length led to the use of geared, manually driven, naval gun mounts and variable-speed, belt-driven, machine gun mounts with the telescopes substituted for the armament In all such systems, the tracking operator observes the missile through an optical sight while controlling the orientation of the telescope to ensure that the missile remains within its field (Fig 4).

The requirements for precise tracking by heavy, long-focal-length telescopes led to the development of complex telescope mounts such as the ROTI Mk II, a 24-in. aperture, 100- to 500-in. focal-length telescope with automatic focus, automatic exposure control, and a 10- to 60-frame/sec Photosonics 70-mm camera for missile photography The 3000-lb telescope, carried by a 9000-lb mount, is driven by an electrohydraulic servo system In the normal aided-tracking mode of operation, the operator observes the missile through a 20-, 30-, or 40-power tracking sight (see Fig. 5). He controls the motion of the telescope by exerting light finger pressure on a control knob. The system must track at rates up to 10°/sec with a precision of 1-2 min of arc if the system resolution of ½ sec is not to be lost because of relative motion of the image during the exposure.

The telescope and mount are shielded from the thermally disturbing sunlight by a dome. The assembled mount is located on a 20-ft tower to prevent degradation of the resolution by the thermally disturbed air immediately over the terrain under the line of sight (Fig. 6).

The precise tracking of modern tracking telescopes has permitted the use of slit spectrographs for infrared and visible-light spectroscopy.

The consistent performance of the 24-in telescopes has led to the preliminary design of 40-in.-aperture telescopes for photographic, spectrographic, and television uses.

The instruments of increased tracking range under development will be general-purpose instruments of 36- to 48-in. aperture with photoelectric or television image detectors mounted on precise mounts with automatic theodolite capability These instruments will supply both engineering data and spatial position for greater portions of the trajectory See ASTRONOMICAL PHOTOGRAPHY; CAMERA, LENS, OPTICAL; SATELLITE, ARTIFICIAL. [C.A.E.]

Optics

Narrowly, the science of light and vision; more broadly, the study of the phenomena associated with the generation, transmission, and detection of electromagnetic radiation in the spectral range extending from the long-wave edge of the x-ray region to the short-wave edge of the radio region This range, often called the optical region or the optical spectrum, extends in wavelength from about 10 angstroms to about 1 mm. Optics has various branches: see METEOROLOGICAL OPTICS; OPTICS, GEOMETRICAL; OPTICS, PHYSICAL; VISION.

In ancient times there was some isolated elementary knowledge of optics, but it was the discoveries of the experimentalists of the early seventeenth century which formed the basis of the science of optics. The statement of the law of refraction by W. Snell, Galileo Galilei's development of the astronomical telescope and his discoveries with it, F. M. Grimaldi's observations of diffraction, and the principles of the propagation of light enunciated by C. Huygens and P de Fermat all came in this relatively short period. The publication of Sir Isaac Newton's *Opticks* in 1704, with its comprehensive and original studies of refraction, dispersion, interference, diffraction, and polarization established the science.

$$\bar{\alpha} = \beta + \gamma \frac{z'}{n'}$$
$$\bar{\beta} = \beta - \alpha \frac{z}{n} + \delta \frac{z'}{n'} - \gamma \frac{zz'}{nn'} \qquad (12)$$
$$\bar{\gamma} = \gamma$$
$$\bar{\delta} = \delta - \gamma \frac{z}{n}$$

The quantity γ, which is independent of the ift (invariant), is called the power of the system. system for which γ is zero is called an afocal stem

Conjugate points Shifting the image origin so at β vanishes leads to

$$x' = \bar{\alpha}x \qquad \xi' = \gamma x + \bar{\delta}\xi$$
$$y' = \bar{\alpha}y \qquad \eta' = \gamma y + \bar{\delta}\eta \qquad (13)$$

The rays from the object origin $x = y = z = 0$ eet at the image origin $x' = y' = z' = 0$ The nage origin thus obtained is said to be conjugate the object origin All the rays from a point $x = x_0, y = y_0$ meet at a point $x'_0 = \bar{\alpha}x_0, y'_0 = \bar{\alpha}y_0$. This means that an object in the plane $z = 0$ is, within the limits of validity of Gaussian optics, maged sharply with a magnification $m = \bar{\alpha}$ in the plane $z' = 0$ In the special case that $m = \bar{\alpha} = 1$, the two conjugate points are called *principal* points. If $m = \bar{\alpha} = n/n'$, the points are called nodal points In case the object is at one of the principal points, it is imaged at the other principal point with unit magnification, in case the ray from ... passes through

origin the distance $z'/n' = -\alpha/\gamma$ makes $\bar{\alpha} = 0$, that is, since $\bar{\beta}\bar{\gamma} = -1$, it follows that

$$x' = -\xi/\gamma \qquad \xi' = \gamma x + \bar{\delta}\xi$$
$$y' = -\eta/\gamma \qquad \eta' = \gamma y + \bar{\delta}\eta \qquad (14)$$

In this case a system of parallel rays (coming from an "infinite axis point," in the language of optics) meets at the image origin This point is called the image focal point A system of parallel rays (direction ξ_0, η_0) is imaged sharply at a point of the focal plane ($x'_0 = -\xi_0/\gamma, y'_0 = -\eta_0/\gamma$)

Shifting the object origin the distance $z/n = \delta/\gamma$ makes $\bar{\delta} = 0$, that is, since $\beta\gamma = -1$, it follows that

$$x' = \bar{\alpha}x - \xi/\gamma \qquad \xi' = \gamma x$$
$$y' = \alpha y - \eta/\gamma \qquad \eta' = \gamma y \qquad (15)$$

The rays from the object origin (which is called the object focal point) emerge parallel to the axis The rays from an arbitrary point x_0, y_0 of the object focal plane $z = 0$ emerge parallel to one another ($\xi' = \gamma x_0, \eta' = \gamma y_0$)

Afocal systems. The power γ is invariant against a shift of object and image origin Until now $\gamma \neq 0$ had to be assumed The case in which $\gamma = 0$ leads to a system which is called afocal, since it has no finite focal point. When object or image is shifted so that $\bar{\beta}$ vanishes the following are obtained, since $\alpha\delta = 1$:

$$x' = \alpha x \qquad \xi' = \xi/\alpha$$
$$y' = \alpha y \qquad \eta' = \eta/\alpha \qquad (16)$$

Any object is imaged with the constant magnification α, and a *parallel bundle of object rays* emerges as a *parallel bundle* with an angular magnification $n'\alpha/n$ for all rays

Types of optical systems It is customary to say that an object-side parallel bundle is a bundle that comes from an object point at infinity, and a parallel emerging bundle is a bundle that is said to be focused at an infinite image point. Therefore, there are four kinds of optical systems corresponding to the four choices of origins just considered. (1) enlarging systems, object and image are at finite conjugates; (2) photographic objectives, a distant object is imaged in the focal plane of the optical system (*see* LENS, OPTICAL); (3) eyepieces and microscope objectives, a near object is imaged at infinity to be seen by the relaxed eye (*see* EYEPIECE, MICROSCOPE, OPTICAL); and (4) telescopes, a distant object is imaged at infinity to be seen by the relaxed eye (*see* TELESCOPE)

In an afocal system, the magnification remains constant for all object distances When the origin is chosen at conjugate points, $\beta = \gamma = 0$, $\delta = 1/\alpha$, and the distances of z and z' of object and image respectively from the origin and the magnification m are given by

$$z'/n' = \alpha^2 z/n$$
$$m = \alpha \qquad (17)$$

On the other hand, in a system of finite power, the magnification changes from object point to object point ... at the two focal points

$$zz' = -nn'\gamma^2$$
$$m = z'\gamma/n' = -n/z\gamma \qquad (18)$$

The distance from the principal point ($m = 1$) to the focal point is given by

$$-z' = n'/\gamma = f'$$
$$-z = n/\gamma = f \qquad (19)$$

The quantity f' is called the focal length of the optical system *See* FOCAL LENGTH

The distance from the nodal point to the focal point is found, by setting $m = n/n'$ in Eqs. (18), to be

$$-z' = -n/\gamma = -f$$
$$z = n'/\gamma = -f' \qquad (20)$$

The nodal points and the principal points coincide for $n = n'$. For a single refracting spherical surface, the nodal points coincide with the center

system is chosen (origin O and x, y, z axes in object space and origin O' and x', y', z' axes in image space), a ray in object space is specified by the coordinates $x, y, 0$ of its intersection point with the plane $z = 0$ and by the optical direction cosines ξ and η formed by the ray with the x and y axes. The ray in image space is specified in the same way with primed coordinates. (An optical direction cosine is a direction cosine multiplied by the refractive index of the respective medium.)

The eight quantities $x, y, x', y', \xi, \eta, \xi'$, and η' are, however, not independent. There exists a characteristic function E of, for instance, x, y, x', and y', from which the other four, ξ, η, ξ', and η', can be computed.

It is found that

$$
\begin{aligned}
-\xi &= \frac{\partial E}{\partial x} = E_x \qquad \xi' = \frac{\partial E}{\partial x'} = E_{x'} \\
-\eta &= \frac{\partial E}{\partial y} = E_y \qquad \eta' = \frac{\partial E}{\partial y'} = E_{y'}
\end{aligned}
\tag{3}
$$

where $E_x, \ldots,$ are introduced as abbreviations for $\partial E/\partial x, \ldots.$

The characteristic function has a physical meaning. It is the optical path—the sum of the paths in each medium multiplied by the corresponding refractive indices—from starting point (coordinates $x, y, 0$) to final point (coordinates $x', y', 0$). The validity of Eq (2) presupposes that x, y, x', and y' determine a single ray, that is, that no two rays from a point in the plane $z = 0$ go through the same point in the plane $z' = 0$.

In case the optical system has an axis of rotation, as do most optical systems, the origins O and O' are best chosen on the axis of rotation, which will be the z (z') axis. Then the x' axis may be chosen parallel to the x axis and in the same direction, and the y' axis parallel to the y axis In this case, the characteristic function depends only on three parameters, for instance

$$
\begin{aligned}
e_1 &= \tfrac{1}{2}(x^2 + y^2) \\
e_2 &= xx' + yy' \\
e_3 &= \tfrac{1}{2}(x'^2 + y'^2)
\end{aligned}
\tag{4}
$$

and Eqs. (3) transform to

$$
\begin{aligned}
-\xi &= E_1 x + E_2 x' \qquad \xi' = E_2 x + E_3 x' \\
-\eta &= E_1 y + E_2 y' \qquad \eta' = E_2 y + E_3 y'
\end{aligned}
\tag{5}
$$

where $E_i = \partial E/\partial e_i$ is introduced as an abbreviation.

When ξ' and η' as well as x' and y' are known, the intersection point of the rays with an arbitrary plane at the distance z' from the image origin can be computed and thus the image formation on any plane or curved surface investigated.

The characteristic function for any special image formation can be given in explicit form. For instance, the characteristic function for imaging the plane $z = 0$ onto a plane at the distance z_0 from the origin with constant magnification but without distortion is

$$
E = n'z'_0 \left[1 + \frac{2}{(m_0 z'_0)^2}(e_1 + m_0 e_2 + m_0^2 e_3)\right]^{1/2} + f(e_1) \tag{6}
$$

where f is an arbitrary function of e_1. The characteristic function for imaging the plane $z = 0$ onto the surface

$$
\begin{aligned}
z' &= \phi(x^2 + y^2) \\
y' &= my \qquad x' = mx
\end{aligned}
\tag{7}
$$

with m and therefore z' being given functions of e_1 leads to

$$
E = n'z' \left[1 + \frac{2}{(mz')^2}(e_1 + me_2 + m^2 e_3)\right]^{1/2} \tag{8}
$$

This leads to a sharp image of the points of a plane with field curvature and distortion present

The existence of the characteristic function can be used to prove that it is impossible to image sharply more than one plane except in a trivial case. The only such image formation possible would be an image comparable to that formed by a plane mirror, in which each object is imaged sharply and undistorted with a magnification which is equal to the ratio n/n' of object and image space See MIRROR OPTICS.

[illegible] only if the [illegible] ler [illegible] ni- fication m_1 and m_2 of the first and of the second surface respectively must obey the condition

$$
m_1 m_2 = n^2/n'^2 \tag{9}
$$

Gaussian optics. The first approximation to optical image formation is called Gaussian optics It describes the rays which are so near the axis that one can assume $x, y, \xi, \eta, x', y', \xi'$, and η' to be so small that only linear terms of their Taylor series need be considered. This gives the position and magnification of the image for small apertures and, if the image is fairly sharp, plane, and undistorted, it also gives, at least approximately, the corresponding data for the image of a finite object with finite field.

The equations

$$
\begin{aligned}
x' &= \alpha x + \beta x' \qquad \xi' = \gamma x + \delta y' \\
y' &= \alpha y + \beta y' \qquad \eta' = \gamma y + \delta y'
\end{aligned}
\tag{10}
$$

describe the image formation if α, β, γ, and δ are assumed to be constant and connected by the relation

$$
\alpha\delta - \beta\gamma = 1 \tag{11}
$$

Equations (10) and (11) give the image coordinates as functions of the object coordinates. Equations of this type, which correspond to the ray-tracing formula, are called direct equations. Shifting the origin on object and image side by the amounts z and z' respectively changes the coefficients α, β, γ, and δ as follows:

$$M = E_2/\zeta' \qquad N = E_3/\zeta'$$

$$\zeta' = n'\left[1 - \frac{2}{n'^2}\left(E_2^2 e_1 + E_2 E_3 e_2 + E_3^2 e_3\right)\right]^{1/2} \quad (26)$$

There is a differential relation between M and N, amely

$$M_3 - N_2 = (2Me_1 + Ne_2)(N_2 M - NM_2) + (Me_2 + 2Ne_3)(MN_3 - NM_3) \quad (27)$$

If E, M, and N are developed in a Taylor series ith respect to the e_i

$$E = \Sigma \bar{E}_i e_i + \tfrac{1}{2}\Sigma \bar{E}_{ik} e_i e_k + \tfrac{1}{6}\Sigma \bar{E}_{ik\lambda} e_i e_k e_\lambda + \cdots$$
$$M = \Sigma \bar{M}_i e_i + \tfrac{1}{2}\Sigma \bar{M}_{ik} e_i e_k + \tfrac{1}{6}\Sigma \bar{M}_{ik\lambda} e_i e_k e_\lambda + \cdots \quad (28)$$
$$N = \Sigma \bar{N}_i e_i + \tfrac{1}{2}\Sigma \bar{N}_{ik} e_i e_k + \tfrac{1}{6}\Sigma \bar{M}_{ik\lambda} e_i e_k e_\lambda + \cdots$$

t can be shown that

$$\begin{aligned} \bar{M}_{ik\lambda} \ldots &= \bar{E}_{2ik\lambda} \ldots + \text{lower-order terms} \\ \bar{N}_{ik\lambda} \ldots &= \bar{E}_{3ik\lambda} \ldots + \text{lower-order terms} \end{aligned} \quad (29)$$

whereby the lower-order terms of $M_{3k\lambda} \ldots$ and $\bar{N}_{2k\lambda} \ldots$ in general will not be the same though hey both start with $\bar{E}_{23k\lambda}$, which has led to controversies about the number of image errors The ndependent image errors are given by the derivaives of E_2 and E_3 Thus, if object and aperture order are considered to be equivalent, that is, if a larger and larger area surrounding the axis is considered, there are five errors of the first order (frequently called third order) given by the so-called error coefficients

$$\bar{E}_{21}, \bar{E}_{22}, \bar{E}_{23}, \bar{E}_{31}, \bar{E}_{33}$$

and nine errors of the second order (frequently called fifth order)

$$\bar{E}_{211}, \bar{E}_{212}, \bar{E}_{213}, \bar{E}_{222}, \bar{E}_{223}, \bar{E}_{233}, \bar{E}_{311}, \bar{E}_{313}, \bar{E}_{333}$$

In short, there are $\binom{n+3}{n+1} - 1$ image-error coefficients of order n corresponding to $2\binom{n+2}{n}$ not necessarily independent image errors of order n, namely

$$\bar{M}_1, \bar{M}_2, \bar{M}_3 \qquad \bar{N}_1, \bar{N}_2, \bar{N}_3$$

for first-order errors and

$$\bar{M}_{11}, \bar{M}_{12}, \bar{M}_{13}, \bar{M}_{22}, \bar{M}_{23}, \bar{M}_{33}$$
$$\bar{N}_{11}, \bar{N}_{12}, \bar{N}_{13}, \bar{N}_{22}, \bar{N}_{23}, \bar{N}_{33}$$

for second-order errors.

When a system with sharp image formation for every point is investigated, it is found that M and N and therefore E_2 and E_3 are functions of e_1 alone

The coefficients $\bar{N}_1, \bar{N}_{11}, \ldots$, or

$$\bar{E}_{31}, \bar{E}_{311}, \bar{E}_{3111}$$

determine the curvature of the image and $\bar{M}_1$, $\bar{M}_{11}, \ldots$, or

$$\bar{E}_{211}, \bar{E}_{2111}, \bar{E}_{21111}, \ldots$$

determine the change of magnification or the errors of distortion. These errors may exist even if every point is sharply imaged.

For an axis point ($x = y = 0$), N becomes a function of e_3 alone Then the rays through a set of concentric circles in the exit pupil go through a set of concentric circles in the image plane, and the image of the object point is concentric. It has been suggested that these errors be called aperture errors The name spherical aberration is used in the literature.

A point on the axis of a cylindrical torus or an ellipsoid has two planes of symmetry. Thus the rays through a set of concentric circles in the exit pupil go through a set of curves with two axes of symmetry and the same center of symmetry in the image plane. The corresponding configuration in the image plane is a set of deformed circles. The corresponding errors may be called deformation errors

The rays from an off axis point have a plane of symmetry, the meridional plane, through the object point and the axis The image in the plane at the distance z' from the exit pupil thus has only one symmetry axis The deviation from double symmetry is caused by the coefficients which have odd powers in e_2 These errors may be called asymmetry or coma errors.

Thus a specific error coefficient may be said to have an order with respect to aperture and field, a field rank and an aperture rank, and a degree of deformation, of coma, or of both For instance, the coefficient

$$E_{11222333}$$

has order 8, field rank 7, aperture rank 11, and coma degree 3 ($o_8 f_7 a_{11} c_3$).

In this case the image-error coefficients (and the corresponding image errors of first order or third order in common practice) can be characterized as

E_{12}	M_1	$o_1 f_3 a_1 c_1$	(Distortion)
E_{13}	N_1	$o_1 f_2 a_2$	(Curvature, sagittal)
E_{22}	M_2	$o_1 f_2 a_2 d_1$	(Astigmatism)
E_{23}	M_3, N_2	$o_1 f_1 a_3 c_1$	(Coma)
E_{33}	N_3	$o_1 f_0 a_4$	(Spherical aberration)

The error usually called astigmatism is in this nomenclature a deformation error; the rays through an aperture circle go through an ellipse in the image plane.

The fifth-order errors can correspondingly be characterized as

E_{112}	M_{11}	$o_2 f_6 a_1$	(Field coefficient of distortion)
E_{113}	N_{11}	$o_2 f_4 a_2$	(Field coefficient of curvature)
$E_{122} \cdot M_{12}$		$o_2 f_4 a_2 d_1$	(Field coefficient of first-order coma)
$E_{122} \cdot M_{12}, N_{11}$		$o_2 f_3 a_3 c_1$	(Field coefficient of aperture error)
$E_{133} \cdot N_{13}$		$o_2 f_2 a_4$	(Field coefficient of aperture error)

of the refracting surface and the principal points coincide with the vertex of the refracting surface.

When the origins are chosen at the principal points ($\bar{\alpha} = \bar{\delta} = 1, \beta = 0$), Eqs. (12) give the following for the distances s and s' of conjugate points and their magnification:

$$\frac{n'}{s'} - \frac{n}{s} = -\gamma = \frac{n'}{f'} = -\frac{n}{f}$$
$$m = 1 - \frac{s'}{f} = \frac{1}{1 - s/f'} = \frac{s'}{s} \qquad (21)$$

When the origins are chosen at the nodal points ($\alpha = n/n'$, $\delta = n'/n$, $\beta = 0$), the distances c, c' of conjugate points and their magnifications are given by

$$\frac{1}{n'c'} - \frac{1}{nc} = -\frac{\gamma}{nn'} = \frac{1}{n'f} = -\frac{1}{nf'} \qquad (22)$$

Off-axis points. Gaussian optics considers only rays near the axis, but it is used as an approximation for the trace of finite rays, especially to compute the amount of light which the optical system transmits. Most optical systems, especially photographic lenses, contain a diaphragm which can be stopped down. In general the diaphragm is the smallest aperture for the object point on the axis, since the other lenses can be made big enough to avoid cutting out any light. However, for off-axis points, the first and last surfaces may cut off some light. They are then said to vignette To obtain an idea of the amount of light going through the system (Fig. 1), one must construct in object (or image) space the Gaussian image of the first lens (a), the last lens (c), and the diaphragm (b), considering not only the position of each but also its magnification

When these three apertures are projected from the object point onto a plane (for instance, the image of the diaphragm in object or image space, called the entrance pupil or the exit pupil), three eccentric circles in the plane of the entrance (exit)

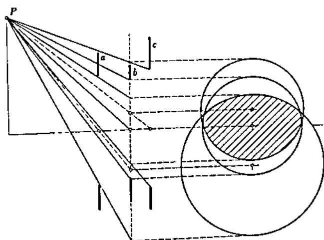


Fig 1. Projection of apertures from finite point P onto entrance pupil (From M Herzberger, Modern Geometrical Optics, Interscience, 1958)

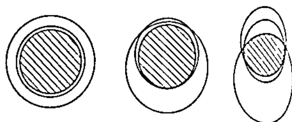


Fig. 2. Change in vignetting diagram as point moves off axis (From M. Herzberger, Modern Geometrical Optics, Interscience, 1958)

pupil are obtained. The rays from the object point through the region common to the three circles give a measure for the vignetting of the light if the object point moves away from the axis (Fig 2).

Image-error theory. For a system with finite aperture and field (beyond the Gaussian domain), all the rays from a given object point do not generally meet at the Gaussian image point. Such a system is said to have image errors.

Let the object and image origins be chosen at the axis point of the object to be imaged and at the axis point of the exit pupil (the Gaussian image of the aperture stop) respectively. Then E, the characteristic path between the two planes $z = 0$ and $z' = 0$, is a function of

$$e_1 = \tfrac{1}{2}(x^2 + y^2)$$
$$e_2 = xx' + yy' \qquad (23)$$
$$e_3 = \tfrac{1}{2}(x'^2 + y'^2)$$

The quantity e_1 depends only on the position of the object point (is a field coordinate) and e_3 depends on the aperture in which the ray intersects the exit pupil (is an aperture coordinate), while e_2 is a mixed coordinate (linear in both field and aperture)

The image errors are then given by two functions M and N having the property that the intersection point (coordinates $\bar{x}'$, $\bar{y}'$) of the rays with a plane at the distance z' from the exit pupil is designated by

$$\bar{x}' = (1 + Nz')x' + Mz'x$$
$$\bar{y}' = (1 + Nz')y' + Mz'y \qquad (24)$$

If M and N are constant and z' is chosen equal to $-1/N_0$, Eqs. (24) give

$$\bar{x}' = -\frac{M_0}{N_0}x$$
$$\bar{y}' = -\frac{M_0}{N_0}y \qquad (25)$$
$$\bar{z}' = -\frac{1}{N_0}$$

That is, the points of the object plane are imaged sharply and without distortion onto the points of the plane at the distance $z' = -1/N_0$ from the exit point The coefficients of the Taylor series expansion of M and N can therefore be considered as image errors.

Functions M and N can be computed from the characteristic function E [illegible] ound to be

objects varying in the number of "waves" per millimeter is imaged.

The variation in amplitude gives a measure of the deterioration of the image with respect to resolution of gratings, and the change of phase gives information about the asymmetry of the image of the sinusoidal test object.

The plot of the sine-wave response as a function of the frequency is regarded as giving information sufficient to compare objectives of similar construction

Diffraction. It is possible to compute diffraction for an off-axis point from geometrical optics If M and N are known, it is possible to compute

$$
\begin{aligned}
E_2 &= M\Big/\left[1 + \frac{2}{n'^2}(M^2 e_1 + MNe_2 + N^2 e_3)\right]^{1/2} \\
E_3 &= N\Big/\left[1 + \frac{2}{n'^2}(M^2 e_1 + MNe_2 + N^2 e_3)\right]^{1/2}
\end{aligned}
\quad (33)
$$

In view of the integrability condition, Eq (27), these equations when integrated give E as a function of e_2 and e_3 and thus give the phase difference at every point of the exit pupil Integration of the exponential e^{iks}, where s is the sum of E and the distance to a fixed point over the exit pupil, enables the (relative) light intensity at the point in question to be computed [M H]

Bibliography: H Chrétien, *Cours de calcul des combinaisons optiques*, 1938; M Herzberger, *Modern Geometrical Optics*, 1958, A. Kerber, *Beiträge zur Dioptrik*, 1896–1899, O Schade, *Optical Image Evaluation*, Nat. Bur Standards Circ. 526, 1954.

Optics, physical

The study of the interaction of electromagnetic waves in the optical range with material systems is called physical optics. The optical range of wavelengths may be taken as the range from about 10 angstroms (10^{-6} mm) to about 1 mm. More narrowly, physical optics deals with the relationship between the atomic structure of a system and the manner in which the system affects light sent into it The chief founder of this branch of science was Michael Faraday, who in 1845 provided the first clue to the electromagnetic nature of light by showing that the optical properties of glass could be altered by a magnetic field (*see* FARADAY EFFECT)

The explanation of the absorption, reflection, scattering polarization, and dispersion of light by a material medium in terms of the properties of the atoms and molecules making up the medium is the objective of physical optics In the course of seeking this objective physicists have found that optical investigations are powerful methods of determining the structures of atoms and molecules and of larger systems composed thereof. *See* ABSORPTION (ELECTROMAGNETIC RADIATION), CRYSTAL OPTICS, DIFFRACTION, DISPERSION (RADIATION), ELECTROMAGNETIC RADIATION; ELECTROOPTICS, FLUORESCENCE; INTERFERENCE OF WAVES; LIGHT, MAGNETOOPTICS; POLARIZED LIGHT; REFLECTION (ELECTROMAGNETIC RADIATION), REFRACTION OF WAVES; SCATTERING (ELECTROMAGNETIC RADIATION); SPECTROSCOPY. *See also* ATOMIC STRUCTURE AND SPECTRA; MOLECULAR STRUCTURE AND SPECTRA. [R.C.L.]

Bibliography: M Born and E. Wolf, *Principles of Optics*, 1959; F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed , 1957; R. W. Wood, *Physical Optics*, 3d ed , 1934

Opuntiales

An order of the plant subclass Dicotyledoneae with a single family (Cactaceae) having 120 genera and perhaps 1700 species, all probably indigenous to America These are unique xerophytic (of dry habitats) plants usually bearing spines or bristles. The stems are modified for water storage and photosynthesis (sugar manufacture). The leaves are usually small and scalelike, rarely well developed, generally early deciduous (dropping off) The flowers are mostly large and brilliantly colored. Sepals, petals, and stamens are indefinite in number. Among the species of cactus are many unusual plants grown in gardens as oddities, and a number are cultivated for their beautiful flowers, such as the Christmas cactus and the night-blooming cereus. The giant cactus (*Cereus giganteus*), or saguaro of Arizona, is the largest of the cacti, occasionally reaching a height of 70 ft *See* CACTUS, *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM. [P D S.]

Oral gland

This gland, located in the mouth, secretes fluids that moisten and lubricate the mouth and food and may initiate digestive activity. Fishes and aquatic amphibians have only solitary mucus-secreting cells in the epithelium of the mouth cavity Multicellular glands first appeared in land animals to keep the mouth moist and make food easier to swallow. These glands occur in definite regions and bear distinctive names. Some glands of terrestrial amphibians have a lubricative secretion, others serve to make the tongue sticky for use in catching insects. Some frogs secrete a serous fluid that contains ptyalin a digestive enzyme The oral glands of reptiles are much the same, but are more distinctly grouped In poisonous snakes and the single poisonous lizard, the Gila monster, certain oral glands of the serous type are modified to form venom. Also many of the lizards have glands that are mixed in character, containing both mucous and serous cells. Oral glands are poorly developed in crocodilians and sea turtles Birds bolt their food, yet grain-eaters have numerous glands, some of which secrete ptyalin

Oral glands in mammals. All mammals, except aquatic forms, are well supplied with oral glands There are numerous small glands, such as the labial glands of the lips, buccal glands of the cheeks, lingual glands of the tongue, and palatine glands of the palate. Besides these, there are larger paired sets in mammals that are quite constant from spe-

$E_{222}:M_{22}$	$o_2f_3a_3d_3$	(Third-order coma)
$E_{223}:N_{22},M_{23}$	$o_3f_2a_4d_1$	(Aperture coefficient of astigmatism)
$E_{233}\cdot N_{23},M_{33}$	$o_2f_1a_5c_1$	(Aperture coefficient of first-order coma)
$E_{331}\cdot N_{33}$	o_3a_6	(Aperture coefficient of aperture error)

If the image errors of a curved object are considered, the error coefficients vary. However, it can be shown that the errors not containing the index 1 are unchanged. They are invariant with respect to curvature of object and image surface. In case they are zero, there can exist an object that is sharply imaged.

Since E gives the coordination for any object and image ray, a knowledge of E must suffice to give the image errors for any object. If the coefficients

$$E_1, E_{11}, E_{111}, \ldots$$

which are the aperture errors of the stop, are added to the image-error coefficients, formulas can be obtained for investigating the image errors that arise when object and stop are moved. Here another division of errors is suitable It is obvious that the errors containing 2,2 and 1,3 have the same rank with respect to aperture and field. Combinations of these errors can designate the degree of skewness of the errors, which can be divided into meridional errors and skew errors of first, second, etc. types. The third- and fifth-order errors are then given by the following:

Zero type:
First order: E_{12}, $2E_{13} + E_{22}$, E_{23}, E_{33}
Second order: E_{112}, $E_{113} + 4E_{122}$, $3E_{123} + 2E_{222}$, $E_{133} + 4E_{223}$

First type:
First order: $E_{22} - E_{12}$
Second order: $E_{113} - E_{122}$, $E_{123} - E_{222}$, $E_{133} - E_{223}$

The coefficients of zero order are the meridional coefficients, which transform by themselves, and the coefficients of each order can then be chosen so that they transform by themselves if the object (or stop) position is changed This means that, for each type, there exists a coefficient which is invariant against the position of both the object and the stop. The first such invariant was found by Josef Petzval and equals $E_{22} - E_{13}$, the Petzval condition; the next would be $E_{2222} - 2E_{2213} + E_{1313}$, and so on

Another analysis of the image errors can be made by considering the diapoint configuration. The diapoint of the object point for a ray is defined as the point where the ray intersects the meridional plane. The coordinates x'_p, y'_p, and z'_p of this point are given by

$$\begin{aligned} x'_p &= -(M/N)x \\ y'_p &= -(M/N)y \\ z'_p &= -1/N \end{aligned} \quad (30)$$

Thus the three-dimensional problem is transformed into a plane problem because one can set $x = 0$ without loss of generality. The coefficient of the development of M/N can be considered as lateral errors and those of $1/N$ as longitudinal errors.

Interpolation theory. In analyzing an optical system, it is not appropriate to develop E into a Taylor series since such a series converges slowly and does not give a good enough approximation for the rays from an object point that is distant from the axis or for a system having a large aperture. It is, however, possible to derive an interpolation formula which gives a very good approximation A number of rays, for instance nine, are traced from a point $x = 0$, $y = y_0$ (for an axis point $y_0 = 0$) through the optical system, and the intersection points $\bar{x}'$, $\bar{y}'$ with a plane at the fixed distance z'_0 from the exit pupil are determined. From

$$\begin{aligned} \bar{x}' &= (1 + Nz'_0)x' \\ \bar{y}' &= (1 + Nz'_0)y' + Mz'_0y_0 \end{aligned} \quad (31)$$

a series of values for M and N as functions of e_2 and e_3 is obtained. These values are fitted by least squares with the formulas

$$\begin{aligned} M &= M_2e_2 + M_3e_3 + \tfrac{1}{2}M_{22}e_2^2 + M_{23}e_2e_3 + \tfrac{1}{2}M_{33}e_3^2 \\ N &= N_2e_2 + N_3e_3 + \tfrac{1}{2}N_{22}e_2^2 + N_{23}e_2e_3 + \tfrac{1}{2}N_{33}e_3^2 \end{aligned} \quad (32)$$

Having found the coefficients of Eqs. (32), it is possible to compute M and N and therefore $\bar{x}'$ and $\bar{y}'$ for a large number of rays from the object point going through the vignetted exit pupil. If the exit pupil is uniformly illuminated, these points should be chosen so that they uniformly fill the (vignetted) exit pupil. The plot of the intersecting points x', y' with the plane at the distance z' then gives a measure of the intensity of the light distribution in the image.

As a measure of the quality of the image, the reciprocal of the radius of the circle that contains a certain percentage of the rays (for instance, 70, 80, or 90%) may be taken The image also can be dissected into its aperture, comatic, and deformation errors with respect to aperture (keeping the object point, that is, the field, constant). This assists the designer in comparing different design stages since the corresponding figures are easy to analyze, in contrast to the complicated image figures.

Moreover, by integrating in the x or y direction, the spread function can be obtained; that is, the image of a line in the meridional direction or the sagittal direction can be investigated.

A simple mathematical consideration shows that a small sinusoidal test object at the object point is imaged sinusoidally but with a different amplitude and phase. (A sinusoidal test object is a pattern in which the intensity varies like a sine wave in the lateral direction, being kept nt in the longitudinal direction.) A usoida

feed; citrus molasses is also used as a livestock feed supplement Since about 1945 an increasing proportion of the orange crop has been used in frozen concentrate which can be marketed the year around In 1955–1956 orange concentrate uti-

(TREE) DISEASES [F E C.]

Orangutan

A primate, Pongo pygmaeus, a member of the ape family Pongidae, found on the islands of Sumatra and Borneo The orangutan is about 4 ft tall, sometimes slightly taller. A large male will weigh 200 lb; females are somewhat smaller. This ape has weak legs but strong, long arms which it uses in swinging through the treetops It is covered with long, loose, reddish brown hair The skin is bluish-gray



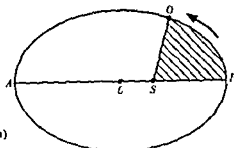
The orangutan, *Pongo pygmaeus*, height to 50 in. (From P. M Duncan, ed, *Cassell's Natural History*, Cassell)

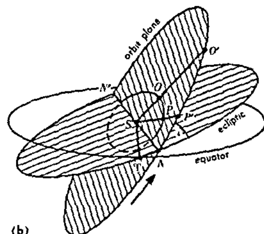
Orangutans live almost entirely in treetops, nesting and sleeping on platforms built in trees. They travel in family groups. Their food is varied, but a favorite item is the fruit of the durian The name orangutan means man-of-the-woods. See PRIMATES. [J.D.B.]

Orbital motion

In astronomy the motion of a material body through space under the influence of its own inertia, a central force, and other forces. Johann Kepler found empirically that the orbital motions of the planets about the Sun are ellipses. Sir Isaac Newton, starting from his laws of motion, proved that an inverse-square gravitational field of force requires a body to move in an orbit that is a circle, ellipse, parabola, or hyperbola.

Elliptical orbit. Two bodies revolving under their mutual gravitational attraction, but otherwise undisturbed, describe orbits of the same shape about a common center of mass. The less massive body has the larger orbit. In the solar system, Sun and Jupiter have a center of mass just outside the visible disk of the Sun For each of the other planets, the center of mass of Sun and planet lies within the Sun.



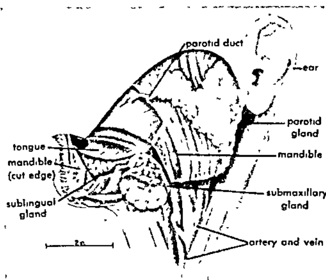


Parameters of an elliptical orbit. (a) The relative orbit. (b) The orbit in space.

For this reason, it is convenient to consider only the relative motion of a planet of mass m about Sun of mass M as though the planet had no mass and moved about a center of mass $M + m$. The orbit so determined is exactly the same shape as the true orbits of planet and Sun about their common center of mass, but it is enlarged in the ratio $(M + m)/M$. See CENTER OF MASS; PLANET.

Parameters of elliptical orbit. The diagram shows the elements or parameters of an elliptic orbit Major axis AP intersects the ellipse AOP at the apsides; the extension of the major axis is the line of apsides. The body is nearest the center of mass at one apside, called perihelion P, and is farthest away at the other, called aphelion A.

Shape and size of an orbit are defined by two elements: length of semimajor axis, and departure of the orbit from a circle Semimajor axis a equals CP; this length is ... distance fr... CS/CP whe... a focus. Fo... ...cal orbits e is always less than unity.



The salivary glands, shown by a partial dissection of the head. (*After J. C. Brash, ed., Cunningham's Textbook of Anatomy, 9th ed, Oxford, 1951*)

cies to species and are commonly designated as salivary glands The parotid gland, near each ear, discharges into the vestibule The submaxillary or submandibular gland lies along the posterior part of the lower jaw; its duct opens well forward under the tongue. The sublingual gland lies in the floor of the mouth. It is really a group of glands, each with its duct. Although not present in man, the retrolingual gland, situated near the submaxillary, is found in many mammals; its duct takes a course similar to that of the submaxillary Other occasional types are the molar gland of the hoofed mammals and the orbital gland of the dog family.

Development. All of the oral glands develop from the epithelial lining as branching buds Each gland is organized somewhat after the pattern of a bush that bears berries on the ends of its twigs. The main stem and all branches of the bush correspond to a system of branching glandular ducts of various sizes; the terminal berries correspond to the secretory end pieces Actually these end pieces are more or less elongate, like the catkins of the willow or birch The ducts are simple epithelial tubes The end pieces specialize in different ways. Some elaborate a serous secretion that typically contains an enzyme; others secrete a mucous fluid, still others contain both types of secretory cells In man and most other mammals the parotid gland produces a purely serous secretion. The submaxillary and sublingual glands of man and most mammals are mixed, or seromucous The secretion of the sublingual gland tends to be more highly mucous in composition than that of the submaxillary

Secretions. Saliva is a viscid fluid containing a mixture of all of the oral secretions. It contains mucus, proteins, salts, and the enzymes ptyalin and maltase. Most of the ptyalin in human saliva is furnished by the parotid gland. The digestive action of saliva is limited to starchy food Other uses of saliva include the moistening of food for easier manipulation by the tongue, the consequent facilitation of swallowing, and a lubrication by mucus that ensures a smoother passage of food down the esophagus to the stomach. The daily amount for man is about 1.5 quarts, for the cow, 65 quarts. *See* GLAND. [L.B.]

Orange

The orange is the most widely used species of citrus fruit and commercially is the most important The sweet orange, *Citrus sinensis*, is a native of China, but it has spread to other tropical and subtropical regions of the world. The sour or bitter oranges, of lesser importance, are quite different from the sweet oranges and belong in a separate species, *C. aurantium*. This article deals only with the sweet orange which includes certain abnormal types such as the navel and blood oranges

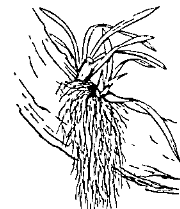
The sweet orange tree is a moderately vigorous evergreen with a rounded, densely foliated top (*see* EVERGREEN PLANTS). The fruits are round or somewhat elongated and orange-colored when ripe. Depending on variety, they may be either seedy or seedless, sometimes with navels or with streaks of red in the flesh. There are many varieties, covering a wide range of ripening time from early to late, thus providing fresh fruit throughout most of the calendar year.

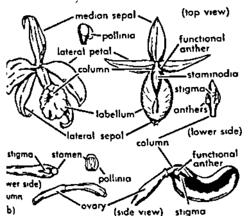
In 1955–1956 the United States produced 38.5% of the world supply of oranges. Florida, with 420,000 acres, leads production, followed by California with 155,900 acres. Texas, Arizona, and Louisiana also grow oranges but with much smaller acreages. From 1948 to 1958 the average annual value of the orange crop of the United States, delivered at the packing house or processing plant, was approximately $223,000,000

Sweet orange fruit is consumed fresh, canned, or as frozen juices; oils from the peel are used in perfumes and flavoring; after the juice is extracted, the rind and flesh are dried and ground for cattle



Foliage, flowers, and fruit of sweet orange, *Citrus sinensis* (*From L. H Bailey, ed, The Standard Cyclopedia of Horticulture, vol* ... *illan, 1937*)





Aerial roots of an orchid epiphytic upon the bark the branch of a tree (*from A J Kerner von Marr-un, The Natural History of Plants, Holt, 1895*) (b) ie two common floral types found in the orchids :presented by *Cattleya* and *Cypripedium* (*F Mc-eel from D B Swingle, A Textbook of Systematic otany, 3d ed., McGraw-Hill, 1946*)

:ollected, dried, and marketed as salep, which is ised both as a medicine and a food *See* ORCHI-DALES; VANILLA [P D S]

Orchidales

An order of the plant class Monocotyledoneae characterized by very minute seeds, each having an undifferentiated embryo and little or no endosperm There are two families the Burmanniaceae or burmannia family, mainly nongreen saprophytes of tropical and subtropical regions, and the Orchidaceae or orchid family, an enormous assemblage of 450 genera with perhaps 15,000 species, widely distributed but most abundant in the tropics where the greatest number are epiphytes The irregular flowers of orchids, developed in connection with insect pollination, are often grotesque but beautiful. Orchids are highly valued as ornamentals of the conservatory, but *Vanilla fragrans*, a Mexican orchid, is the source of commercial vanilla *See* ORCHID; VANILLA, *see also* EMBRYOPHYTA, MONOCOTYLEDONEAE, PLANT KINGDOM. [P.D.S.]

Ordovician

The second period of the Paleozoic Era, and the system of rocks deposited during this time—the succession of rocks overlying the Cambrian system and underlying the Silurian system The Ordovician period had a duration of some 60,000,000 to 80,000,000 years.

PRE-CAMBRIAN		PALEOZOIC								MESOZOIC			CENOZOIC	
						CARBONIFEROUS								
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

... named by C. Lapworth, Ordovices, an , of Wales before the coming of the Romans. The system included parts of the original Cambrian system of A. Sedgwick and of the Silurian system of R Murchison, specifically those strata which succeeded the Tremadoc and underlay the Landovery beds, and were characterized by distinctive fossil graptolites This nomenclature is followed in most of the world, but in some countries it is the practice to continue the Silurian in its original extended sense, assigning the Lower Silurian or Untersilur to the Ordovician, and the Upper Silurian to the Gotlandian system, named from a Swedish island in the Baltic Sea

Rocks. The typical Ordovician rocks of Wales are extremely variable but generally consist of thousands of feet of graywacke and argillite with associated lavas and volcanic fragmental rocks. The sequence has been divided into several series; in ascending order they are the Arenigian, Llanvirnian, Llandeilan, Caradocian, and Ashgillian. Each is composed of one or more fossil zones, characterized by assemblages of graptolites. The sequence of graptolite zones established in Britain has been found to be generally adaptable to rocks ...

LITHINA

In North America, rocks having the characteristic Ordovician graptolites are found in many localities from northeast Newfoundland to eastern Tennessee and southeastern Oklahoma along the Atlantic and Gulf coasts, and from southeastern Alaska and eastern British Columbia to central Idaho and central Nevada along the Pacific side of the continent. They are in sequences of graywackes,

Position of a body in its orbit at time t can be computed if a, e, and time of perihelion passage p and period of revolution T are known. Let O be the position of a planet at time t and OSP be the area swept out in time $t - p$ (*see* AREAL VELOCITY). From Kepler's area law, area OSP equals $(t - p)/T$ multiplied by the area of the full ellipse.

To describe the orientation of an orbit in space, several other parameters are required. All orbits in the solar system are referred to the plane of the ecliptic, this being the plane of the orbit of Earth about the Sun. The reference point for measurement of celestial longitude in the plane of the ecliptic is the vernal equinox ♈, the First Point of Aries. This is the point where the apparent path of the Sun crosses the Earth's Equator from south to north. The two points of intersection of the orbit plane with the plane of the ecliptic (N and N') are called the nodes, and the line joining them is the line of nodes Ascending node N is the one where the planet crosses the plane of the ecliptic in a northward direction, N' is the descending node. The angle as seen from Sun S measured in the plane of the ecliptic from the vernal equinox to the ascending node is ♈SN, it is termed the longitude of the ascending node ☊, and fixes the orbit plane with respect to the zero point of longitude The angle at the ascending node between the plane of the ecliptic and the orbit plane is called the inclination i and defines the orientation of the orbit plane with respect to the fundamental plane. The angle as seen from the Sun, measured in the orbit plane from the ascending node to perihelion, is NSP' and is referred to as the argument of perihelion, it defines the orientation of the ellipse within the orbit plane. The angle (NSP' + ☊), measured in two different planes, is called the longitude of perihelion $\bar{\omega}$ Because dynamically the semimajor axis a and period T of a planet of mass m revolving under influence of gravitation G about Sun of mass M are related by the expression

$$\frac{4\pi^2}{T^2} = \frac{G(M+m)}{a^3}$$

only six elements, a, e, i, ☊, $\bar{\omega}$, and p, are required to fix the position of a planet in space Instead of these elements, however, a position vector (x,y,z) and the associated velocity vector $(\dot{x},\dot{y},\dot{z})$ at a given instant of time would serve equally well to define the path of a planet in a rectangular coordinate system with origin at the Sun.

Orbital velocity. Orbital velocity v of a planet moving in a relative orbit about the Sun may be expressed by

$$v^2 = G(M+m)\left(\frac{2}{r} - \frac{1}{a}\right)$$

where a is the semimajor axis, and r is the distance from the planet to the Sun. In the special case of a circular orbit, $r = a$, and the expression becomes

$$v^2 = \frac{G(M+m)}{a}$$

When the eccentricity of an orbit is exactly unity the length of the major axis becomes infinite and the ellipse degenerates into a parabola The expression for the velocity then becomes

$$v^2 = G(M+m)\left(\frac{2}{r}\right)$$

This parabolic velocity is referred to as the velocity of escape, it being the minimum velocity required for a particle to escape from the gravitational attraction of its parent body (*see* ESCAPE VELOCITY). Eccentricities greater than unity occur with hyperbolic orbits Because in a hyperbola the semimajor axis a is negative, hyperbolic velocities are greater than the escape velocity.

Parabolic and hyperbolic velocities seem to be observed in the motions of some comets and meteors. Aside from the periodic ones, most comets appear to be visitors from cosmic distances, as do about two-thirds of the fainter meteors. For ease of computation, the short arcs of these orbits that are observed near perihelion are represented by parabolas rather than ellipses. Although the observed deviation from parabolic motion is not sufficient to vitiate this computational procedure, it is possible that many of these "parabolic" comets are actually moving in elliptical orbits of extremely long period The close approach of one of these visitors to a massive planet, such as Jupiter, could change the velocity from parabolic to elliptical if retarded, or from parabolic to hyperbolic if accelerated It is possible that many of the periodic comets, especially those with periods under nine years, have been captured in this way. *See* AREAL VELOCITY; CELESTIAL MECHANICS, COMET; GRAVITATION; PERTURBATION (ASTRONOMY).

[R.L.D]

Bibliography: P Herget, *The Computation of Orbits*, 1948, F R Moulton, *Introduction to Celestial Mechanics*, 2d ed, 1948, W S. Smart, *Spherical Astronomy*, 1931.

Orchid

Any member of the orchid family (Orchidaceae), one of the largest families of plants, with 450 genera and perhaps 15,000 species. Orchids have a wide distribution, being most abundant in tropical forests where the majority are epiphytes (live perched on other plants). In the temperate and arctic regions, the genera are terrestrial. Unusual features of this group are the complex and highly specialized flowers, the minute microscopic seeds having no endosperm, and the great number of seeds in an orchid capsule. The extraordinary beauty of the flowers makes orchids the basis of a multimillion-dollar floral industry. Otherwise, these plants have little economic importance. Vanilla is obtained from the pods of *Vanilla planifolia*, and the tubers of some Asiatic species are

nents The Ordovician rocks yield the first great variety of conodonts, forms that seem quite useful in classifying some rocks that are otherwise sparsely fossiliferous Sponge fossils are occasionally common and distinctive; and there are representatives of other fossil invertebrate classes and orders The advent of the first vertebrate may have preceded the Ordovician, but at least there are scales of primitive fishes in some abundance in rocks of about Blackriveran age, particularly in the Harding sandstone of Colorado Calcareous algae are the only plant fossils of consequence Of all the Ordovician animal life, graptolites are probably the most distinctive, for though they lived from Cambrian to Carboniferous, they are almost entirely limited to the Ordovician and Silurian systems

Tectonic provinces. The Ordovician rocks of North America fall into several tectonic provinces (Fig 2). The central part of the continent or hedreocraton accumulated a few hundred feet to a thousand feet or so of sedimentary rocks, principally limestone and dolomite, it was a relatively stable area. Along the eastern and western borders of the craton, separated by zones of crustal flexure, were belts having much thicker sections (a mile or more) of sed [illegible]

bonat [illegible]

preva [illegible]

greater subsidence than the hedreocraton Beyond them to the edge of the continent, Ordovician rocks are almost entirely terrigenous and volcanic, consisting of graywackes and argillites with associated lava flows and fragmental volcanic rocks, these belts of subsidence had associated islands that rose and were eroded, as well as volcanic centers, and are eugeosynclinal belts. The rocks in these belts were greatly deformed by later mountain making, and in many cases were so invaded by plutonic [illegible]neous rocks and were so metamorphosed as to be [di]fficult to identify and date, or to arrange in strati[gr]aphic order. The typical Ordovician of Great [Bri]tain is eugeosynclinal, as is that of coastal [S]candinavia; whereas that of Sweden, Estonia, and [P]oland is similar to that of interior north America [S]ee GEOSYNCLINE

Deformation in North America. The most striking structural changes in the continent were along the eastern margin, from Newfoundland to the South Atlantic Coast Initially, in the Canadian epoch, graptolite-bearing rocks of eugeosynclinal facies were laid in submerged troughs adjoining islands along the coast, while carbonates, which were thick in the miogeosynclinal belt, thinned toward the cratonal interior area The interrelations of the two facies are obscured by later folding and thrust faulting along the zone of change, the carbonate rocks were laid in shallow water and probably graded into the argillites laid in deeper marginal troughs. In later epochs, at different times along the 1000-mile length of the belt in eastern United States, lands rose in the eugeosynclinal belt and shed sediment into troughs extending into the margin of the hedreocraton. The effect of the rising lands became pronounced in the Cincinnatian epoch, when sands and muds spread inland as far as Ohio and Michigan in a great delta that filled an elongate basin centered in Pennsylvania to a depth of several thousand feet. The later Ordovician rocks were folded, and Silurian rocks lie unconformably on them at localities from Gaspé, Quebec, to Pennsylvania. Moreover, there are great thrust faults attributed to this, the Taconian orogeny, and intrusions of granitic rocks in maritime Canada and New England that were unroofed (exposed by erosion) by Silurian time; intrusions in the Carolinas have also been dated as Ordovician by geochemical methods. *See* GEOCHRONOMETRY; UNCONFORMITY

In western North America, there is similar contrast between volcanic-bearing argillaceous and graywacke sequences with graptolites as the most frequent fossils from southeastern Alaska and Yukon to central Nevada and the southern Sierra Nevada of California and carbonate rocks of a mile or so thickness extending eastward to the cratonal margin. Though the zone of contact between the facies is again obscure, in some areas the two facies are known to grade into each other, as though the carbonate rocks were laid in shallow water passing over a flexure into deeper sinking troughs that received terrigenous sediment from lands raised in the eugeosynclinal belt to the west. Orogeny is not recognized in the west during this period

In the continental interior, carbonate rocks are prevalent, and in the Late Ordovician (late Trentonian and the Cincinnatian) seas covered all but a very small part of the continental interior from the Gulf of Mexico to the Arctic. During the Late Ordovician there were islands and volcanoes along the present borders of North America, but the

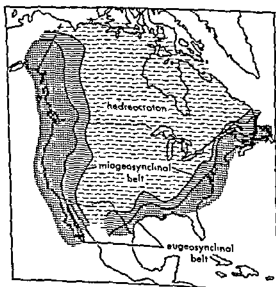


Fig 2 North American Ordovician tectonic provinces.

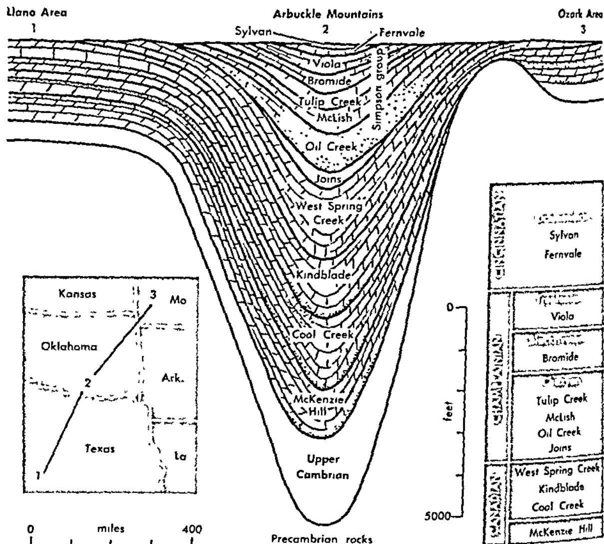


Fig 1. Ordovician deposits in the Llano, Arbuckle, and Ozark areas. Diagram emphasizes greater thickness and completeness of sedimentary record in geosynclinal area of southern Oklahoma. *(Adapted fr... R. C Moore, Introduction to Historical Geolog... McGraw-Hill, 2d ed., 1958)*

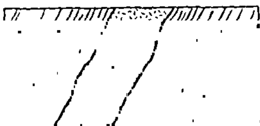
argillites, and volcanic rocks as in Wales. On the other hand, graptolites are few or lacking in carbonate rocks and associated shales and sandstones that are classified as Ordovician in the interior of the continent; however, these are placed in the Ordovician system because graptolites are occasionally found in association with the other fossils and some of the carbonate rocks can be traced stratigraphically into graptolite-bearing shales.

The North American classification is based on the rocks of shelly facies, in which brachiopods and cephalopods are generally the most definitive fossils. The rocks have been divided into series ... ascending order: Canadian ...

Mohawkian is used frequently for the Blackriveran and Trentonian, and Bolarian is a provincial series approximating the Blackriveran. Champlainian has been used for the whole system, and more recently it has been used for the three middle series as listed above. The base of the Carado...

riveran, and the Ashgillian is about equivalent ... the Cincinnatian.

Life. The Ordovician contrasts with the older a... underlying Cambrian in having a greater abundan... and variety of fossils. Trilobites and brachiopo... are common in many Cambrian rocks, but oth... organisms are rarely found In the Ordovician roc... of shaly facies, graptolites are generally abunda... only on occasional beds But in the calcareo... rocks or shelly facies, fossils abound in ma... places. Brachiopods are perhaps the most general... present; they are of great variety and differ fro... stage to stage. Though less frequent, cephalopo... and trilobites are quite useful in recognition ... ages of beds, particularly in the Canadian an... Chazyan series. Corals became plentiful enough ... form small patch reefs in the Chazyan; and bry... zoans first appeared in some abundance in th... series. Well-developed crinoids and cystids a... common in a few limited zones. Gastropods an... pelecypods are sometimes abundant, particularly i... argillaceous rocks, and ostracods are first know...


Fig 1. Vein developed in fissured or sheeted zone

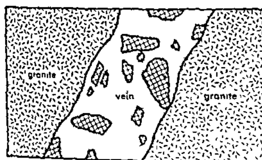


Fig 2 Brecciated vein in granite

ock along a fissure (Fig 1) Replacement bodies n limestone may be very irregular Veins are usually inclined steeply and may either cut across or conform with the bedding or foliation of the enclosing rocks The inclination is called the dip, and is the angle between the vein and the horizontal. The horizontal trend of the vein is its *strike*, and the vertical angle between a horizontal plane and the line of maximum elongation of the vein is the plunge. Commonly the veins of a mining district occur as systems which have a general strike, and one or more systems may be present at some angle to the main series In places the mineralization is a network of small, irregular, discontinuous veins called a stockwork

Mineral deposits are seldom equally rich throughout The pay ore may occur in streaks, spots, bunches, or bands separated by low-grade material or by gangue These concentrations of valuable ore are called ore shoots, if roughly horizontal they are ore horizons, and if steeply inclined they are chimneys After their formation mineral deposits may be deformed by folding, faulting, or brecciation (Fig 2)

Metasomatism or replacement Metasomatism, or replacement, is the process of essentially simultaneous removal of one mineral and deposition in its place of another mineral of partly or wholly different composition. A large volume of rock may be transformed in this manner, and the resulting deposit is generally of equal volume. Commonly the original structure and texture of the replaced rock is preserved by the replacing material

Replacement, evidence for which is found in many mineral deposits, operates at all depths under a wide range of temperature The evidence indicates that the new minerals formed in response to conditions that were unstable for the preexisting ones

Usually the replacing material moves to the site of metasomatism along relatively large openings such as faults, fractures, bedding planes, and shear zones. It then penetrates the rock along smaller cracks and finally enters individual mineral grains along cleavage planes and minute fractures where substitution may take place on an atomic scale until the entire mass has been transformed (Fig. 3). In many deposits repeated movement has opened and reopened channelways, which would otherwise have become clogged, to permit continued and widespread replacement. The process may take place through the action of gases or solutions or by reactions in the solid state.

Classification of mineral deposits. Mineral deposits are generally classified on the basis of the geologic processes responsible for their formation as magmatic, contact metasomatic, pegmatitic, hydrothermal, sedimentary, residual, and regional

the concentrated minerals form part of the body of the igneous rock. If the magma solidified by simple crystallization, the economically valuable mineral is distributed through the resulting rock; diamond deposits found in peridotite are believed by some geologists to be of this type. However, if the magma has differentiated during crystallization, early-formed minerals may settle to the bottom of the magma chamber and form segregations such as the chromite deposits of the Bushveld in South Africa Late-formed minerals may crystallize in the interstices of older minerals and form segregations like the Bushveld platinum deposits. Occasionally, the residual magma becomes enriched in constituents such as iron, and this enriched liquid may form deposits, such as the Taberg titaniferous iron ores of Sweden It is also possible that during differentiation some of the crystals or liquid may be injected and form sills or dikes The iron ores of Kiruna, Sweden, have been described as early injections, and certain pegmatites are classed as late magmatic injections. Magmatic deposits are relatively simple in mineral composition and few in number.

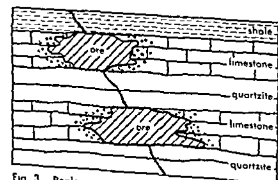


Fig 3 Replacement of limestone by ore along fissure Disseminated ore (dots) is forming in advance of main body

interior was largely beneath the sea. The interior, however, was not fully stable, for the thicknesses of rocks in the stages of the several series thicken into basins of greater subsidence and thin to disappearance along the margins of other areas that subsided little or not at all. At one stage in the Chazyan Epoch, sands from dunes in the northern interior drifted into shallow seas in the Mississippi Valley region, forming the few hundred feet of remarkably pure St. Peter sandstone, the source of silica for glass manufacture and other chemical industries.

Other continents. Ordovician rocks are known from each of the continents, but knowledge of them comes principally from Europe and North America. In South America, Ordovician fossils have been described from areas scattered from Venezuela and Columbia to northwestern Argentina along the west slope of the Andes. In Asia, the Ordovician of Manchuria has been a source of faunas, but little is known of the paleogeography of the continent. In Australia, the best-known Ordovician is that in Victoria, having an excellent graptolite succession, and similar faunas are known in New Zealand. The principal Paleozoic sections in Africa are in the countries bordering the western Mediterranean.

[M.K.]

Bibliography: W. H. Twenhofel et al., Correlation of the Ordovician formations of North America, *Bull Geol. Soc. Am.*, 65(3):247–298, 1954.

Ore and mineral deposits

Ore deposits are naturally occurring geologic bodies that may be worked for one or more metals. The metals may be present as native elements, or, more commonly, as oxides, sulfides, sulfates, silicates, or other compounds The term ore is often used loosely to include such nonmetallic minerals as fluorite and gypsum The broader term, mineral deposits, includes, in addition to metalliferous minerals, any other useful minerals or rocks Minerals of little or no value which occur with ore minerals are called gangue. Some gangue minerals may not be worthless in that they are used as by-products, for instance, limestone for fertilizer or flux, pyrite for making sulfuric acid, and rock for road material.

Mineral deposits that are essentially as originally formed are called primary or hypogene. The term hypogene also indicates formation by upward movement of material. Deposits that have been altered by weathering or other superficial processes are secondary or supergene deposits Mineral deposits that formed at the same time as the enclosing rock are called syngenetic, and those formed later are called epigenetic.

The distinction between metallic and nonmetallic deposits is at times an arbitrary one since some substances classified as nonmetals, such as lepidolite, spodumene, beryl, and rhodochrosite, are the source of metals. The principal reasons for distinguishing nonmetallic deposits from metallic are practical ones, and include such economic considerations as methods of recovery and uses

Concentration. The earth's crust consists of igneous, sedimentary, and metamorphic rocks Table 1 gives the essential composition of the crust and shows that 10 elements make up more than 99% of the total Of these, aluminum, iron, and magnesium are industrial metals. The other metals are present in small quantities, mostly in igneous rocks (*Table 2*)

Many mineral deposits, such as stone and salt, are mined in the condition in which they formed without further concentration. However, most deposits are natural enrichments and concentrations of original material produced by different geologic processes. To be of commercial grade, the metals must be present in much higher concentrations than the averages shown in Table 2. For example, the following metals must be concentrated in the amounts indicated to be considered ores: aluminum, about 30%, copper, 0.7–10%, lead, 2–4%; zinc, 3–8%; and gold, silver, and uranium, only a small fraction of a per cent of metal Therefore, natural processes of concentration have increased the aluminum content of aluminum ore 3–4 times, and even a low grade gold ore may represent a concentration of 20,000 times

Forms of mineral deposits. Mineral deposits occur in many forms depending upon their origin, later deformation, and changes caused by weathering. Syngenetic deposits are generally sheetlike, tabular, or lenticular, but may on occasion be irregular or roughly spherical.

Epigenetic deposits exhibit a variety of forms. Veins or lodes are tabular or sheetlike bodies that originate by filling fissures [illegible] the country

Table 1. Elemental composition of earth's crust based on igneous and sedimentary rocks*

	Weight, %	Atom, %	Volume, %
Oxygen	46.71	60.5	91.24
Silicon	27.69	20.5	0.51
Titanium	0.62	0.3	0.03
Aluminum	8.07	6.2	0.44
Iron	5.05	1.9	0.37
Magnesium	2.08	1.8	0.28
Calcium	3.65	1.9	1.04
Sodium	2.75	2.5	1.21
Potassium	2.58	1.4	1.88
Hydrogen	0.14	3.0	

* From T. F. W Barth, *Theoretical Petrology*, Wiley, 1952 (recalculated from F. W. Clarke and H S Washing-

Table 2. Abundance of metals in igneous rocks

Element	%	Element	%
Aluminum	8.13	Cobalt	0.0023
Iron	5.00	Lead	0.0016
Magnesium	2.09	Arsenic	0.0005
Titanium	0.44	Uranium	0.0004
Manganese	0.10	Molybdenum	0.00025
Chromium	0.02	Tungsten	0.00015
Vanadium	0.015	Antimony	0.0001
Zinc	0.011	Mercury	0.00005
Nickel	0.008	Silver	0.00001
Copper	0.005	Gold	0.0000005
Tin	0.004	Platinum	0.0000005

Laboratory synthesis of sulfide minerals by G Kullerud shows that some ore-bearing solutions must have been considerably more concentrated than is generally believed *See* SULFIDE PHASE EQUILIBRIA; *see also* ORE DEPOSITS, GEOCHEMISTRY OF.

Two common features of hydrothermal deposits are the zonal arrangement of minerals and alteration of wall rock.

1. *Zoning of mineralization.* Many ore deposits change in composition with depth, lateral distance, or both, resulting in a zonal arrangement of minerals or elements This arrangement is generally interpreted as being due to deposition from solution with decreasing temperature and pressure, the solution precipitating minerals in reverse order of their solubilities. Other factors are also involved such as concentration, relative abundance, decrease electrode potentials, and reactions within the itions and with the wall rocks as precipitation gresses

'onal distribution of minerals was first noted in ieral deposits associated in space with large eous bodies, and has since been extended to inde zoning related to sedimentary and metamorc processes in places where no igneous bodies in evidence Although most geologists interpret ing as a result of precipitation from a single ending solution, some believe deposition is ieved from solutions of different ages and of different compositions

The distribution of mineral zones is clearly wn at Cornwall, England and at Butte, Mont At rnwall

rd ti

ser,

e carbonates Such zoning is by no means a unisal phenomenon, and, in addition to mines and tricts where it is lacking, there are places where ersals of zones occur. Some of these reversals ve been explained more or less satisfactorily by scoping of minerals near the surface, by the ects of structural control or of composition of host rock in precipitating certain minerals, and the effects of supergene enrichment on the original zoning but many discrepancies are not adeiately explained

2. *Wall rock alteration* The wall rocks of hyothermal deposits are generally altered, the most mmon change being a bleaching and softening here alteration has been intense, as in many esothermal deposits, primary textures may be iliterated by the alteration products Chemical id mineralogical changes occur as a result of the troduction of some elements and the removal of hers; rarely a rearrangement of minerals takes ace with no replacement.

Common alteration products of epithermal and esothermal deposits are quartz, sericite, clay minals, chlorite, carbonates, and pyrite. Under highmperature hypogene conditions pyroxene, amhibole, biotite, garnet, topaz, and tourmaline form many mines sericite has been developed nearest e vein and gives way outward to clay minerals or chlorite. The nature and intensity of alteration vary with size of the vein, character of the wall rock, and temperature and pressure of hydrothermal fluids. In the large, low-grade porphyry copper and molybdenum deposits associated with stocklike intrusives, alteration is intense and widespread, and two or more stages of alteration may be superimposed.

Under low-intensity conditions, the nature of the wall rock to a large extent determines the alteration product. High-intensity conditions, however, may result in similar alteration products regardless of the nature of the original rock. Exceptions to this are monomineralic rocks such as sandstones and limestones Wall rock alteration may develop during more than one period by fluids of differing compositions, or it may form during one period of mineralization as the result of the action of hydrothermal fluids that did not change markedly in composition Alteration zones have been used as guides to ore and tend to be most useful where they are neither too extensive nor too narrow. Mapping of these zones outlines the mineralized area and may indicate favorable places for exploration

Sedimentary and residual deposits. At the earth's surface, action of the atmosphere and hydrosphere alters minerals and forms new ones that are more stable under the existing conditions Sedimentary deposits are bedded deposits derived from preexisting material by weathering, erosion, transportation, deposition, and consolidation Different source materials and variations in the processes of formation yield different deposits. Changes that take place in a sediment after it has formed and before the succeeding material is laid down are termed diagenetic. They include compaction, solution, recrystallization, and replacement (*see* DIAGENESIS). In general, the sediment is consolidated by compaction and by precipitation of material as a cement between mineral grains.

The mineral deposits that form as a result of sedimentary and weathering processes are commonly grouped as follows: (1) sedimentary deposits, not including products of evaporation, (2) chemical evaporites, (3) placer deposits (4)

dolomite beds, sulfur deposits such as those near Kuibyshev in Russia and the deposits of the Gulf Coast region, and the phosphate of North Africa and Florida Metalliferous deposits such as the minette iron ores of Lorraine and Luxemborg, the Clinton iron ores of the United States, and the manganese of Tchiaturi, Georgia, and Nikopol in the Ukraine also belong here. There are other deposits of metals in sedimentary rocks whose origin remains an enigma, such as the uranium of the Colorado Plateau, the Witwatersrand in South Africa, and Blind River in Ontario, and the copper deposits of Mansfeld, Germany, and of the Copper-belt of Northern Rhodesia and the Belgian Congo.

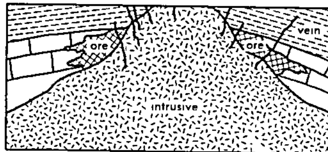


Fig 4. Association of contact metasomatic and vein deposits with intrusive.

Contact metasomatic deposits. During the crystallization of certain magmas a considerable amount of fluid escapes. This fluid may produce widespread changes near the contacts of magma with the surrounding rocks (Fig. 4). Where such changes are caused by heat effects, without addition of material from the magma, the resulting deposits are called contact metamorphic. If appreciable material is contributed by the magma, the deposits are termed contact metasomatic. The magmas that produce these effects are largely silicic in composition and the resulting mineral deposits are often irregular in form.

Under contact metasomatic conditions, the introduced fluids extensively replace the country rock to produce a variety of complex minerals Contact metasomatic deposits include a number of important deposits, whereas contact metamorphic deposits are rarely of economic value Many garnet, emery, and graphite deposits are classed as contact metasomatic, as are such metalliferous deposits as the iron ores of Cornwall, Pa., Iron Springs, Utah, and Banat, Hungary; many copper ores of Utah, Arizona, New Mexico, and Mexico, the zinc ores of Hanover, N.M.; and various tungsten ores of California and Nevada.

Pegmatite deposits. Pegmatites are relatively coarse-grained rocks found in igneous and metamorphic regions. The great majority of them consist of feldspar and quartz, often accompanied by mica, but complex pegmatites contain unusual minerals and rare elements. Many pegmatites are regular tabular bodies; others are highly irregular and grade into the surrounding rocks In size, pegmatites range from a few inches in length to bodies over 1000 ft long and scores of feet across. Some pegmatites are zoned, commonly with a core of quartz surrounded by zones in which one or two minerals predominate.

Pegmatites may originate by various igneous and metamorphic processes. Fractional crystallization of a magma results in residual solutions that are generally rich in alkalies, alumina, water, and other volatiles. The volatiles lower the temperature of this liquid and make it unusually fluid; the low viscosity promotes the formation of coarse-grained minerals. The rare elements that were unable by substitution to enter into the crystal structure of earlier-formed minerals, principally because of differences in size of their atomic radii, are concentrated in the residual pegmatite solutions Late hydrothermal fluids may alter some of the previously formed pegmatite minerals.

Some pegmatites develop by replacement of the country rock and commonly these are isolated bodies with no feeders or channels in depth. They occur in metamorphic regions usually devoid of igneous rocks and contain essentially the same minerals as those in the country rocks In some regions small pegmatites have grown by forcing apart the surrounding metamorphic rock, and others have formed by filling a fissure or crack from the walls inward. In both cases growth is believed to have taken place by diffusion and consolidation of material in the solid state.

Hydrothermal deposits. Most vein and replacement deposits are believed to be the result of precipitation of mineral matter from dilute, hot ascending fluids. As the temperature and pressure decrease, deposition of dissolved material takes place There is no general agreement as to the state of these fluids. Some geologists believe that they were gaseous emanations that condensed to hot solutions, whereas others think they began as solutions and remained so until precipitation occurred

W. Lindgren, who developed the hydrothermal theory, divided these deposits into three groups on the basis of temperature and pressure conditions supposed to exist at the time of formation. Deposits formed at temperatures of 50–200°C and at slight depth beneath the surface are called epithermal Many ores of mercury, antimony, gold, and silver are of this type. Deposits formed between 200 and 300°C at moderate depths are known as mesothermal and include ores of gold-quartz, silver lead, copper, and numerous other types. Hypothermal deposits are those formed between about 300 and 500°C at high pressures; certain tin, tungsten, and gold-quartz ores belong to this type.

The nature of hydrothermal fluids is determined by inference, by analogy with laboratory experiments, and by investigation of deposits forming around volcanoes and hot springs at the present time. Studies of liquid inclusions in minerals, of mineral textures, and of inversion temperatures of minerals indicate that mineralization takes place at elevated temperatures. Layers of minerals on the walls of open fissures with crystal faces developed toward the openings suggest deposition from solution. In some of these cavities later crystals were deposited on earlier ones in a manner that suggests growth in moving solutions Certain secondary replacement phenomena, such as weathering and oxidation of mineral deposits, also indicate deposition from liquid solutions. Studies of wall rock alteration where hydrothermal solutions have attacked and replaced rock minerals indicate that these solutions change in character from place to place

The principal objections to the hydrothermal theory are the low solubility of sulfides in water and the enormous quantities of water required W. Lindgren realized this and, for some deposits, favored colloidal solutions as carriers of metals

Laboratory synthesis of sulfide minerals by G Kullerud shows that some ore bearing solutions must have been considerably more concentrated than is generally believed *See* SULFIDE PHASE EQUILIBRIA; *see also* ORE DEPOSITS, GEOCHEMISTRY OF.

Two common features of hydrothermal deposits are the zonal arrangement of minerals and alteration of wall rock.

1 Zoning of mineralization. Many ore deposits change in composition with depth, lateral distance, or both, resulting in a zonal arrangement of minerals or elements This arrangement is generally interpreted as being due to deposition from solution with decreasing temperature and pressure, the solution precipitating minerals in reverse order of their solubilities. Other factors are also involved such as concentration, relative abundance decrease in electrode potentials, and reactions within the solutions and with the wall rocks as precipitation progresses

Zonal distribution of minerals was first noted in mineral deposits associated in space with large igneous bodies, and has since been extended to include zoning related to sedimentary and metamorphic processes in places where no igneous bodies are in evidence. Although most geologists interpret zoning as a result of precipitation from a single ascending solution, some believe deposition is achieved from solutions of different ages and of different compositions

The distribution of mineral zones is clearly shown at Cornwall, England, and at Butte, Mont At Cornwall, tin veins in depth pass upward and outward into copper veins, followed by veins of lead silver, then antimony, and finally iron and manganese carbonates Such zoning is by no means a universal phenomenon, and, in addition to mines and districts where it is lacking, there are places where reversals of zones occur Some of these reversals have been explained more or less satisfactorily by telescoping of minerals near the surface, by the effects of structural control or of composition of the host rock in precipitating certain minerals and by the effects of supergene enrichment on the original zoning, but many discrepancies are not adequately explained

2. Wall rock alteration The wall rocks of hydrothermal deposits are generally altered, the most common change being a bleaching and softening Where alteration has been intense, as in many mesothermal deposits, primary textures may be obliterated by the alteration products Chemical and mineralogical changes occur as a result of the introduction of some elements and the removal of others; rarely a rearrangement of minerals takes place with no replacement

Common alteration products of epithermal and mesothermal deposits are quartz, sericite, clay minerals, chlorite, carbonates, and pyrite. Under high-temperature hypogene conditions pyroxene, amphibole, biotite, garnet, topaz, and tourmaline form In many mines sericite has been developed nearest the vein and gives way outward to clay minerals or chlorite. The nature and intensity of alteration vary with size of the vein, character of the wall rock, and temperature and pressure of hydrothermal fluids In the large, low-grade porphyry copper and molybdenum deposits associated with stocklike intrusives, alteration is intense and widespread, and two or more stages of alteration may be superimposed

Under low-intensity conditions, the nature of the wall rock to a large extent determines the alteration product High-intensity conditions, however, may result in similar alteration products regardless of the nature of the original rock. Exceptions to this are monomineralic rocks such as sandstones and limestones Wall rock alteration may develop during more than one period by fluids of differing compositions, or it may form during one period of mineralization as the result of the action of hydrothermal fluids that did not change markedly in composition Alteration zones have been used as guides to ore and tend to be most useful where they are neither too extensive nor too narrow Mapping of these zones outlines the mineralized area and may indicate favorable places for exploration.

Sedimentary and residual deposits At the earth's surface, action of the atmosphere and hydrosphere alters minerals and forms new ones that are more stable under the existing conditions Sedimentary deposits are bedded deposits derived from preexisting material by weathering, erosion, transportation, deposition, and consolidation. Different source materials and variations in the processes of formation yield different deposits Changes that take place in a sediment after it has formed and before the succeeding material is laid down are termed diagenetic. They [illegible]
lization, an [illegible]
general, the [illegible]
tion and by precipitation of material as a cement between mineral grains.

The mineral deposits that form as a result of sedimentary and weathering processes are commonly grouped as follows: (1) sedimentary deposits, not including products of evaporation, (2) chemical evaporites, (3) placer deposits, (4) re- [illegible]

[illegible] this group
[illegible], the great petroleum resources, clay deposits, limestone and dolomite beds, sulfur deposits such as those near Kuibyshev in Russia and the deposits of the Gulf Coast region, and the phosphate of North Africa and Florida Metalliferous deposits such as the minette iron ores of Lorraine and Luxemborg, the Clinton iron ores of the United States, and the manganese of Tchiaturi, Georgia, and Nikopol in the Ukraine also belong here. There are other deposits of metals in sedimentary rocks whose origin remains an enigma, such as the uranium of the Colorado Plateau, the Witwatersrand in South Africa, and Blind River in Ontario; and [illegible]

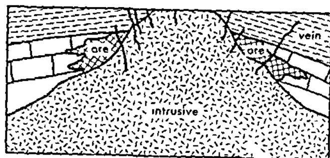


Fig. 4. Association of contact metasomatic and vein deposits with intrusive.

Contact metasomatic deposits During the crystallization of certain magmas a considerable amount of fluid escapes. This fluid may produce widespread changes near the contacts of magma with the surrounding rocks (Fig. 4). Where such changes are caused by heat effects, without addition of material from the magma, the resulting deposits are called contact metamorphic. If appreciable material is contributed by the magma, the deposits are termed contact metasomatic. The magmas that produce these effects are largely silicic in composition and the resulting mineral deposits are often irregular in form.

Under contact metasomatic conditions, the introduced fluids extensively replace the country rock to produce a variety of complex minerals. Contact metasomatic deposits include a number of important deposits, whereas contact metamorphic deposits are rarely of economic value Many garnet, emery, and graphite deposits are classed as contact metasomatic, as are such metalliferous deposits as the iron ores of Cornwall, Pa., Iron Springs, Utah, and Banat, Hungary; many copper ores of Utah, Arizona, New Mexico, and Mexico, the zinc ores of Hanover, N.M ; and various tungsten ores of California and Nevada

Pegmatite deposits Pegmatites are relatively coarse-grained rocks found in igneous and metamorphic regions. The great majority of them consist of feldspar and quartz, often accompanied by mica, but complex pegmatites contain unusual minerals and rare elements. Many pegmatites are regular tabular bodies; others are highly irregular and grade into the surrounding rocks In size, pegmatites range from a few inches in length to bodies over 1000 ft long and scores of feet across. Some pegmatites are zoned, commonly with a core of quartz surrounded by zones in which one or two minerals predominate.

Pegmatites may originate by various igneous and metamorphic processes. Fractional crystallization of a magma results in residual solutions that are generally rich in alkalies, alumina, water, and other volatiles. The volatiles lower the temperature of this liquid and make it unusually fluid; the low viscosity promotes the formation of coarse-grained minerals. The rare elements that were unable by substitution to enter into the crystal structure of earlier-formed minerals, principally because of differences in size of their atomic radii, are concentrated in the residual pegmatite solutions. Late hydrothermal fluids may alter some of the previously formed pegmatite minerals.

Some pegmatites develop by replacement of the country rock and commonly these are isolated bodies with no feeders or channels in depth. They occur in metamorphic regions usually devoid of igneous rocks and contain essentially the same minerals as those in the country rocks. In some regions small pegmatites have grown by forcing apart the surrounding metamorphic rock, and others have formed by filling a fissure or crack from the walls inward. In both cases growth is believed to have taken place by diffusion and consolidation of material in the solid state.

Hydrothermal deposits. Most vein and replacement deposits are believed to be the result of precipitation of mineral matter from dilute, hot ascending fluids. As the temperature and pressure decrease, deposition of dissolved material takes place. There is no general agreement as to the state of these fluids Some geologists believe that they were gaseous emanations that condensed to hot solutions, whereas others think they began as solutions and remained so until precipitation occurred

W Lindgren, who developed the hydrothermal theory, divided these deposits into three groups on the basis of temperature and pressure conditions supposed to exist at the time of formation Deposits formed at temperatures of 50–200°C and at slight depth beneath the surface are called epithermal. Many ores of mercury, antimony, gold, and silver are of this type. Deposits formed between 200 and 300°C at moderate depths are known as mesothermal and include ores of gold-quartz, silver lead, copper, and numerous other types. Hypothermal deposits are those formed between about 300 and 500°C at high pressures; certain tin, tungsten, and gold-quartz ores belong to this type.

The nature of hydrothermal fluids is determined by inference, by analogy with laboratory experiments, and by investigation of deposits forming around volcanoes and hot springs at the present time Studies of liquid inclusions in minerals, of mineral textures, and of inversion temperatures of minerals indicate that mineralization takes place at elevated temperatures Layers of minerals on the walls of open fissures with crystal faces developed toward the openings suggest deposition from solution. In some of these cavities later crystals were deposited on earlier ones in a manner that suggests growth in moving solutions Certain secondary replacement phenomena, such as weathering and oxidation of mineral deposits, also indicate deposition from liquid solutions. Studies of wall rock alteration where hydrothermal solutions have attacked and replaced rock minerals indicate that these solutions change in character from place to place.

The principal objections to the hydrothermal theory are the low solubility of sulfides in water and the enormous quantities of water required W. Lindgren realized this and, for some deposits, favored colloidal solutions as carriers o als.

Laboratory synthesis of sulfide minerals by G Kullerud shows that some ore-bearing solutions must have been considerably more concentrated than is generally believed *See* SULFIDE PHASE EQUILIBRIA; *see also* ORE DEPOSITS, GEOCHEMISTRY OF.

Two common features of hydrothermal deposits are the zonal arrangement of minerals and alteration of wall rock.

1 Zoning of mineralization. Many ore deposits change in composition with depth, lateral distance, or both, resulting in a zonal arrangement of minerals or elements This arrangement is generally interpreted as being due to deposition from solution with decreasing temperature and pressure, the solution precipitating minerals in reverse order of their solubilities. Other factors are also involved such as concentration, relative abundance, decrease in electrode potentials, and reactions within the solutions and with the wall rocks as precipitation progresses

Zonal distribution of minerals was first noted in mineral deposits associated in space with large igneous bodies, and has since been extended to include zoning related to sedimentary and metamorphic processes in places where no igneous bodies are in evidence Although most geologists interpret zoning as a result of precipitation from a single ascending solution, some believe deposition is achieved from solutions of different ages and of different compositions

The distribution of mineral zones is clearly shown at C[illegible]

[illegible] then antimony and [illegible]
nese carbo[illegible]
versal phe[illegible]
districts where it is lacking there are places where reversals of zones occur Some of these reversals have been explained more or less satisfactorily by telescoping of minerals near the surface, by the effects of structural control or of composition of the host rock in precipitating certain minerals and by the effects of supergene enrichment on the original zoning, but many discrepancies are not adequately explained

2 Wall rock alteration The wall rocks of hydrothermal deposits are generally altered, the most common change being a bleaching and softening Where alteration has been intense, as in many mesothermal deposits, primary textures may be obliterated by the alteration products Chemical and mineralogical changes occur as a result of the introduction of some elements and the removal of others; rarely a rearrangement of minerals takes place with no replacement.

Common alteration products of epithermal and mesothermal deposits are quartz, sericite, clay minerals, chlorite, carbonates, and pyrite Under high-temperature hypogene conditions pyroxene, amphibole, biotite, garnet, topaz, and tourmaline form In many mines sericite has been developed nearest the vein and gives way outward to clay minerals or chlorite. The nature and intensity of alteration vary with size of the vein, character of the wall rock, and temperature and pressure of hydrothermal fluids. In the large, low grade porphyry copper and molybdenum deposits associated with stocklike intrusives, alteration is intense and widespread, and two or more stages of alteration may be superimposed

Under low-intensity conditions, the nature of the wall rock to a large extent determines the alteration product. High-intensity conditions, however, may result in similar alteration products regardless of the nature of the original rock. Exceptions to this are monomineralic rocks such as sandstones and limestones Wall rock alteration may develop during more than one period by fluids of differing compositions, or it may form during one period of mineralization as the result of the action of hydrothermal fluids that did not change markedly in composition Alteration zones have been used as guides to ore and tend to be most useful where they are neither too extensive nor too narrow. Mapping of these zones outlines the mineralized area and may indicate favorable places f[illegible]

su[illegible]
alters minerals and forms new ones that are more stable under the existing conditions. Sedimentary deposits are bedded deposits derived from preexisting material by weathering, erosion, transportation, deposition and consolidation Different source materials and variations in the processes of formation yield different deposits. Changes that take place in a sediment after it has formed and before the succeeding material is laid down are termed diagenetic They include compaction, solution, recrystallization, and replacement (*see* DIAGENESIS). In general, the sediment is consolidated by compaction and by precipitation of material as a cement between mineral grains

The mineral deposits that form as a result of sedimentary and weathering processes are commonly grouped as follows. (1) sedimentary deposits, not including products of evaporation, (2) chemical evaporites, (3) placer deposits, (4) re[illegible]

[illegible] limestone and dolomite beds, sulfur deposits such as those near Kuibyshev in Russia and the deposits of the Gulf Coast region, and the phosphate of North Africa and Florida Metalliferous deposits such as the minette iron ores of Lorraine and Luxemborg, the Clinton iron ores of the United States, and the manganese of Tchiaturi, Georgia, and Nikopol in the Ukraine also belong here. There are other deposits of metals in sedimentary rocks whose origin remains an enigma, such as the uranium of the Colorado Plateau, the Witwatersrand in South Africa, and Blind River in Ontario, and [illegible]

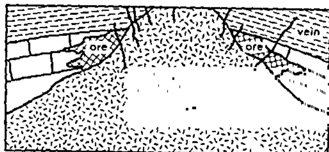


Fig. 4. Association of contact metasomatic and vein deposits with intrusive.

Contact metasomatic deposits. During the crystallization of certain magmas a considerable amount of fluid escapes. This fluid may produce widespread changes near the contacts of magma with the surrounding rocks (Fig. 4). Where such changes are caused by heat effects, without addition of material from the magma, the resulting deposits are called contact metamorphic. If appreciable material is contributed by the magma, the deposits are termed contact metasomatic. The magmas that produce these effects are largely silicic in composition and the resulting mineral deposits are often irregular in form.

Under contact metasomatic conditions, the introduced fluids extensively replace the country rock to produce a variety of complex minerals. Contact metasomatic deposits include a number of important deposits, whereas contact metamorphic deposits are rarely of economic value. Many garnet, emery, and graphite deposits are classed as contact metasomatic, as are such metalliferous deposits as the iron ores of Cornwall, Pa., Iron Springs, Utah, and Banat, Hungary, many copper ores of Utah, Arizona, New Mexico, and Mexico; the zinc ores of Hanover, N.M.; and various tungsten ores of California and Nevada.

Pegmatite deposits. Pegmatites are relatively coarse-grained rocks found in igneous and metamorphic regions. The great majority of them consist of feldspar and quartz, often accompanied by [illegible]

grade into the surrounding rocks. In size, pegmatites range from a few inches in length to bodies over 1000 ft long and scores of feet across. Some pegmatites are zoned, commonly with a core of quartz surrounded by zones in which one or two minerals predominate.

Pegmatites may originate by various igneous and metamorphic processes. Fractional crystallization of a magma results in residual solutions that are generally rich in alkalies, alumina, water, and other volatiles. The volatiles lower the temperature of this liquid and make it unusually fluid; the low viscosity promotes the formation of coarse-grained minerals. The rare elements that were unable by substitution to enter into the crystal structure of earlier-formed minerals, principally because of differences in size of their atomic radii, are concentrated in the residual pegmatite solutions. Late hydrothermal fluids may alter some of the previously formed pegmatite minerals.

Some pegmatites develop by replacement of the country rock and commonly these are isolated bodies with no feeders or channels in depth. They occur in metamorphic regions usually devoid of igneous rocks and contain essentially the same minerals as those in the country rocks. In some regions small pegmatites have grown by forcing apart the surrounding metamorphic rock, and others have formed by filling a fissure or crack from the walls inward. In both cases growth is believed to have taken place by diffusion and consolidation of material in the solid state.

Hydrothermal deposits. Most vein and replacement deposits are believed to be the result of precipitation of mineral matter from dilute, hot ascending fluids. As the temperature and pressure decrease, deposition of dissolved material takes place. There is no general agreement as to the state of these fluids. Some geologists believe that they were gaseous emanations that condensed to hot solutions, whereas others think they began as solutions and remained so until precipitation occurred.

W. Lindgren, who developed the hydrothermal theory, divided these deposits into three groups on the basis of temperature and pressure conditions supposed to exist at the time of formation. Deposits formed at temperatures of 50–200°C and at slight depth beneath the surface are called epithermal. Many ores of mercury, antimony, gold, and silver are of this type. Deposits formed between 200 and 300°C at moderate depths are known as mesothermal and include ores of gold-quartz, silver-lead, copper, and numerous other types. Hypothermal deposits are those formed between about 300 and 500°C at high pressures; certain tin, tungsten, and gold-quartz ores belong to this type.

The nature of hydrothermal fluids is determined by inference, by analogy with laboratory experiments, and by investigation of deposits forming around volcanoes and hot springs at the present time. Studies of liquid inclusions in minerals, of mineral textures, and of inversion temperatures of minerals indicate that mineralization takes place at elevated temperatures. Layers of minerals on the walls of open fissures with crystal faces developed toward the openings suggest deposition from solution. In some of these cavities later crystals were deposited on earlier ones in a manner that suggests growth in moving solutions. Certain secondary replacement phenomena, such as weathering and oxidation of mineral deposits, also indicate deposition from liquid solutions. Studies of wall rock alteration where hydrothermal solutions have attacked and replaced rock minerals indicate that these solutions change in character from place to place.

The principal objections to the hydrothermal theory are the low solubility of sulfides in water and the enormous quantities of water required. W. Lindgren realized this and, for some deposits, favored colloidal solutions as carriers of metals

of minerals, rather than a concentration, would result from the processes operative. However, if movement of material were confined to specific channelways, this objection would not necessarily hold

Oxidation and supergene enrichment. Many sulfide minerals form at depth under conditions differing markedly from those existing at the surface When such minerals are exposed by erosion or deformation to surface or near-surface conditions, they become unstable and break down to form new minerals Essentially all minerals are affected.

The oxidation of mineral deposits is a complex process Some minerals are dissolved completely or in part, whereas elements of others recombine and form new minerals. The principal chemical processes that take place are oxidation, hydration, and *carbonation*. The oxidation of pyrite and other sulfides produces sulfuric acid, a strong solvent Much of the iron in the sulfides is dissolved and reprecipitated as hydroxide to form iron-stained outcrops called gossans Metal and sulfate ions are leached from sulfides and carried downward to be precipitated by the oxidizing waters as concentrations of oxidized ores above the water table. Oxides and carbonates of copper, lead, and zinc form, as do native copper, silver, and gold. The nature of the ore depends upon the composition of the primary minerals and the ...

... ucing agents they are precipitated as secondary sulfides. The oxidized zone may thus pass downward into the supergene sulfide zone Where this process has operated extensively, a thick secondary or supergene enriched sulfide zone is formed Enrichment may take place by removal of valueless material or by solution of valuable metals which are then transported and reprecipitated. This enrichment process has converted many low-grade ore bodies into workable deposits Supergene enrichment is characteristic of copper deposits but may also take place in deposits of other metals Beneath the enriched zone is the primary sulfide ore (Fig 6).

The textures of the gossan minerals may give a clue to the identity of the minerals that existed before oxidation and enrichment took place These have been used as guides in prospecting for ore

Sequence of deposition. Studies of the relations of minerals in time and space have shown that a fairly constant sequence of deposition, or paragenesis, is characteristic of many mineral deposits. In magmatic and contact metasomatic deposits, silicates form first, followed by oxides and then sulfides W. Lindgren presented this sequence for hypogene mineral associations, and A B Edwards has discussed the problems involved The sequence of common minerals starts with quartz, followed by iron sulfides or arsenides, chalcopyrite sphalerite, bornite, tetrahedrite, galena, and complex lead and silver sulfo salts. It has been established primarily by laboratory investigations and indicates the existence of some fundamental control, but attempts to explain the series and variations in it have been largely unsuccessful. Local variations are to be expected since many factors such as replacement, unmixing, superimposed periods of mineralization, structural and stratigraphic factors, and telescoping of minerals may complicate the order of deposition

Paragenesis is generally thought to be the result of decreasing solubility of minerals with decreasing temperature and pressure. It has also been explained in terms of relative solubilities, pH of the solutions, metal volatilities, decreasing order of potentials of elements, free energies, and changing crystal structures of the minerals as they are deposited In order to explain mineral paragenesis more satisfactorily, many additional experimental studies must be made to determine phase relations at different temperatures and pressures *See* MINERAL.

Mineralogenetic provinces and epochs. Mineral deposits are not uniformly distributed in the earth's crust nor did they all form at the same time In certain regions conditions were *favorable for the* concentration of useful minerals. These regions are termed mineralogenetic provinces and they contain broadly similar types of deposits, or deposits with different mineral assemblages that appear to be genetically related The time during which these deposits formed constitutes a mineralogenetic epoch and such epochs differ in duration, but in general they cover a long time interval that is not sharply defined Certain provinces contain mineral deposits of more than one epoch.

During diastrophic periods in the earth's history mountains were formed accompanied by plutonic ...

... supergene enrichment, and mechanical action. The relationship between mineral deposition and large-scale crustal movements permits a grouping of mineralogenetic provinces by major tectonic features of the continents such as mountain belts, sta-

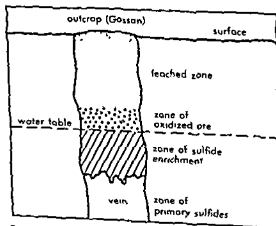


Fig 6 Vein deposit showing changes due to oxidation and supergene enrichment

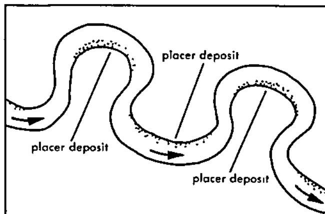


Fig. 5. Deposition of stream placer on inside of meander bends.

These deposits have characteristics of both syngenetic and epigenetic types. A controversy centers around the genesis of these and similar deposits of the world.

2. Chemical evaporites Chemical evaporites consist of soluble salts formed by evaporation in closed or partly closed shallow basins. Deposits of salt or gypsum that are several hundred feet thick are difficult to explain satisfactorily. Oschsenius suggested that they formed in basins which were separated from the ocean by submerged bars except for a narrow channel (inlet); such barriers are common along coastal areas. Intermittently, sea water flowed over the barrier and was concentrated into saline deposits by evaporation Modifications of this theory have been proposed to account for the omissions of certain minerals and the interruptions in the succession.

Deposits of gypsum and common salt (halite) are found in many countries, whereas the larger concentrations of potash salts, borates, and nitrates are much more restricted in occurrence *See* EVAPORITE (SALINE).

3 Placer deposits. Placers are the result of mechanical concentration whereby heavy, chemically resistant, tough minerals are separated by gravity from light, friable minerals. Separation and concentration may be accomplished by streams, waves and currents, air, or by soil and hill creep The most important economic placer deposits are those formed by stream action (Fig. 5).

Stream and beach placers are widespread in occurrence and include the famous gold placers of the world, as well as deposits of magnetite, ilmenite, rutile, zircon, monazite, and garnet Placer deposits of diamond, platinum, and gemstones are less common.

4. Residual deposits Complete weathering results in distribution of the rock as a unit and the segregation of its mineral constituents. This is accomplished by oxidation, hydration, and solution and may be accelerated by the presence of sulfuric acid Some iron and manganese deposits form by accumulation without change, but certain clay and bauxite deposits are created during the weathering of aluminous rocks. Residual concentrations form where relief is not great and where the crust is stable; this permits the accumulation of material in place without erosion. *See* WEATHERING PROCESSES.

Large residual deposits of clay, bauxite, phosphate, iron, and manganese have been worked in many parts of the world, as have smaller deposits of nickel, ocher, and other minerals.

5. Organic deposits. Plants and animals collect and use various inorganic substances in their life processes, and concentration of certain of these substances upon the death of the organisms may result in the formation of a mineral deposit. Coal and peat form from terrestrial plant remains and represent concentration by plants of carbon from the carbon dioxide of the atmosphere. Petroleum originates by the accumulation of plant and animal remains. Many limestone, phosphate, and silica deposits also form by plant and animal activity Hydrated ferric oxide and manganese dioxide are precipitated by microorganisms; anaerobic bacteria can reduce sulfates to sulfur and hydrogen sulfide. There is considerable controversy, however, as to whether microorganisms are responsible for the formation of certain iron, manganese, and sulfide deposits Some uranium, vanadium, copper, and other metalliferous deposits are considered to have formed, in part at least, by the activity of organisms.

Deposits formed by regional metamorphism. Regional metamorphism includes the reconstruction that takes place in rocks within orogenic or mountain belts as a result of changes in temperature, pressure, and chemical environment. In these orogenic belts, rocks are intensely folded, faulted, and subjected to increases in temperature. The changes that occur in this environment affect the chemical and physical stability of minerals, and new minerals, textures, and structures are produced, generally accompanied by the introduction of considerable material and the removal of other material

Some geologists believe that the water and metals released during regional metamorphism can give rise to hydrothermal mineral deposits Along faults and shear zones movement of fluids could take place by mechanical flow, though elsewhere movement might be by diffusion. The elements released from the minerals would migrate to low pressure zones such as brecciated or fissured areas and concentrate into mineral deposits. It has been suggested that the subtraction of certain elements during metamorphism also can result in a relative enrichment in the remaining elements; if this process is sufficiently effective, a mineral deposit may result Certain minerals also may be concentrated during deformation by flow of material to areas of low pressure such as along the crests of folds

Deposits of magnetite, titaniferous iron, and various sulfides may form in metamorphic rocks as well as deposits of nonmetallic minerals such as kyanite, corundum, talc, graphite, and garnet.

Opponents of the concept of mineral formation by regional metamorph[illegible] e that a dispersal

ances, and whether such *rejuvenated and regenerated ores* would be dispersed or concentrated by *the processes* envisaged by Schneiderhohn

[A F H.]

Bibliography: A. M. Bateman, *Economic Mineral Deposits*, 2d ed, 1950; A M Bateman (ed.), *Fiftieth Anniversary Volume, 1905–1955, Economic Geology*, 1955; A. B. Edwards, *Textures of the Ore Minerals*, 2d ed, 1954, G Kullerud et al., Ore minerals, in *Annual Report of the Director of the Geophysical Laboratory*, 1957; W Lindgren, *Mineral Deposits*, 4th ed, 1933, H. E. McKinstry, *Mining Geology*, 1948

Ore deposits, geochemistry of

Geochemistry in general deals with the amounts and distribution of the elements and isotopes of the earth, and the nature of the processes affecting them. Although ore bodies are formed by many different processes (*see* ORE AND MINERAL DEPOSITS), only those deposits of the heavy metals believed to be formed by precipitation from heated water-rich fluids (hydrothermal) are discussed in detail here. Most of the world's supply of base metals, silver, and gold originates in such deposits

Mineral and chemical composition. The minerals in ore deposits frequently are divided into two groups, ore and gangue, the former constitute those for which the deposit is mined, the latter are the waste minerals associated with the ore. The same mineral may be an ore in some deposits and a gangue in others The common minerals of hydrothermal deposits (Table 1) are sulfides, sulfo salts, oxides, carbonates, silicates, and native elements, although sulfates, a fluoride, tungstates, arsenides, tellurides, selenides, and others are by no means rare Many minor elements which seldom occur in sufficient abundance to form discrete minerals of their own may substitute for the major elements of the ore minerals and thus be recovered as by-products For example (as shown in Table 1), the ore mineral of cadmium, indium, and gallium is sphalerite, the major ore mineral of silver and thallium is galena, and pyrite is sometimes an ore of cobalt *See* ELEMENTS (GEOCHEMICAL DISTRIBUTION)

Ore deposits consist, in essence, of exceptional concentrations of given elements over that commonly occurring in rocks The degree of concentration needed to constitute ore varies widely, as shown in Table 2, and is a complex function of many economic and sometimes political variables The quantity of these elements in the total known or reasonably expected ore bodies in the world is infinitesimal when compared with the total amounts in the crust of the earth Thus, each and every cubic mile of ordinary rocks in the crust of the earth contains enough of each ore element to make large bodies (see Table 2) Although there is a large number of geologic situations that are apparently favorable, only a very few of them contain significant amounts of ore Thus it is evident that the processes leading to concentration must be the exception and not the rule, and obviously any understanding or knowledge of these processes should aid in the discovery of further deposits

It is apparent from the above that each step in the process of ore formation must be examined carefully if this sporadic occurrence of ore is to be placed on a rational basis. In order for ores to form, there must be a source for the metal, a medium in which it may be transported, a driving force to move this medium, a "plumbing system" through which it may move, and a cause of precipitation of the ore elements as an ore body These interrelated requirements are discussed below in terms of the origin of the hydrothermal fluid, its chemical properties, and the mechanisms by which it may carry and deposit ore elements.

Source of metals. It is not easy to determine the source for the metals in hydrothermal ore deposits because, as shown above, they exist everywhere in such quantities that even highly inefficient processes could be adequate to extract enough material to form large deposits.

Fluids associated with igneous intrusion In many deposits there is evidence that ore formation was related to the intrusion of igneous rocks nearby, but in many other deposits intensive search has failed to reveal any such association Because the crystal structures of the bulk of the minerals (mostly silicates) crystallizing in igneous rocks are such that the common ore elements such as [illegible] ele- [illegible] uids, [illegible] nces [illegible] r the bulk of the magma has crystallized, are the hydrothermal fluids which move outward and upward to areas of lower pressure in the surrounding rocks, where part or all of their contained metals are precipitated as ores A more detailed discussion of the composition of these fluids is presented below.

Fluids obtained from diagenetic and metamorphic processes Fluids of composition similar to the above also could be obtained from diagenetic and metamorphic processes. When porous, water-saturated sediments containing the usual amounts of hydrous and carbonate minerals are transformed into essentially nonhydrous, nonporous metamorphic rocks, great quantities of water and carbon dioxide must be driven off. Thus, each cubic mile of average shale must lose about 3×10^9 tons of water and may lose large amounts of carbon dioxide on metamorphism to gneiss. The great bulk of the water presumably comes off as connate water (entrapped at time of rock deposition) under conditions of fairly low temperature In many respects, this water has the same sea water composition as it had to start with. However, as metamorphism proceeds, accompanied by slow thermal buildup from heat flow from the earth's interior and from radioactivity, the last fluids are given off at higher temperatures and are richer in CO_2 and other substances These fluids would have considerably greater solvent power and can be expected to

ble regions, and Precambrian shields. *See* TECTONIC PATTERNS.

The Precambrian shield areas of the world contain the Lake Superior, Kiruna, and Venezuelan iron provinces, the gold provinces of Kirkland Lake and Porcupine in Canada, the gold-uranium ores of South Africa, the gold deposits of western Australia, and the base metals of central Australia. In the more stable regions are the metalliferous lead-zinc province of the Mississippi Valley and provinces of salt and gypsum, iron, coal, and petroleum in different parts of the world. The mountain belts are the location of many diverse kinds of mineral provinces such as the gold-quartz provinces of the Coast Range and the Sierra Nevadas, various silver-lead-zinc provinces of the western United States, the Andes, and elsewhere, and numerous base-metal provinces in the Americas, Africa, Australia, and Europe.

Localization of mineral deposits. The foregoing discussion has shown that mineral deposits are localized by geologic features in various regions and at different times. Within the shield areas and mountain belts major mineralized districts are often localized in the upper parts of elongate plutonic bodies (*see* PLUTON). Specific ores tend to occur in particular kinds of rocks. Thus tin, tungsten, and molybdenum are found in granitic rocks, and nickel, chromite, and platinum occur in basic igneous rocks. Tropical climates favor the formation of residual manganese and bauxite deposits, whereas arid and semiarid climates favor the development of thick zones of supergene copper ores. Major mineralized districts are also localized by structural features such as faults, folds, contacts, and intersections of superimposed orogenic belts The location of individual deposits is commonly controlled by structural features, by the physical or chemical characteristics of the host rock (Fig 7), by topographic features, by ground-water action, or by restriction to certain favorable beds (Fig. 8).

Source and transport of ores. Widely divergent views have been expressed as to the original source and mode of transport of mineral deposits, but in general two ideas have predominated for many years According to one view, the source was a differentiating magma which split off fractions as the pressure and temperature changed. Some of the heavy metals crystallized within the magmatic body

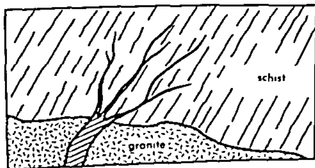


Fig. 7. Strong vein in granite dividing into stringers upon entering schist.

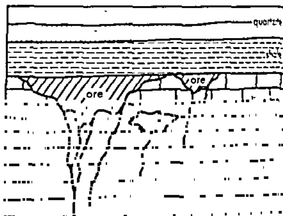


Fig. 8 Ore in limestone beneath impervious shale

and, if sufficiently concentrated, formed true magmatic deposits. Others combined with mineralizing fluids and rose into cooler rocks where they were precipitated as hydrothermal deposits. Thus the magma was both the source and the transporting agent of the ore material. A few geologists believe the source of the ore to be at even greater depths than that at which magmas form.

According to the second view, the materials in mineral deposits were derived from the surrounding rocks in either of two ways:

1. Circulation of surface waters removed metals from the host rocks and deposited them in available openings; this is the lateral secretion theory. The metals were carried either by cool surface waters or by such waters that moved downward, became heated by contact with hot rocks at depth, and then rose and deposited their dissolved material.

2. During regional metamorphism large quantities of hydrothermal fluids may be released from rocks in deep orogenic zones. These fluids remove metals and other minerals from the country rock and redeposit them at higher levels along favorable structures Elements may also move by diffusion along chemical, thermal, and pressure gradients.

A number of the famous mineralized districts of the world that have characteristics of both epigenetic and syngenetic deposits have been modified by later metamorphism, thereby further obscuring their origin In some of these districts the fissure and joint systems in the rocks reflect the pattern in deeper-seated rocks. H Schneiderhohn (Stuttgart) has suggested that repeated rejuvenation of these systems by tectonic movements, accompanied by the dissolving action of thermal waters on old ore deposits in depth, would result in upward movement and reprecipitation of metals in higher formations; Schneiderhohn calls these deposits secondary hydrothermal ores Elsewhere old folded rocks and ore deposits have been greatly deformed and the ores taken into solution and transported to higher and younger strata, such deposits Schneiderhohn terms regenerated ores. Controversy centers around suitable criteria for epigenetic and syngenetic deposits, the problems of solubility of metals in thermal waters, the [illegible] rt over lon [illegible]

tances, and whether such rejuvenated and regenerated ores would be dispersed or concentrated by the processes envisaged by Schneiderhohn.

[A.F.H.]

Bibliography: A. M. Bateman, *Economic Mineral Deposits*, 2d ed., 1950; A. M. Bateman (ed.), *Fiftieth Anniversary Volume, 1905–1955, Economic Geology*, 1955; A. B. Edwards, *Textures of the Ore Minerals*, 2d ed., 1954; G. Kullerud et al., Ore minerals, in *Annual Report of the Director of the Geophysical Laboratory*, 1957; W. Lindgren, *Mineral Deposits*, 4th ed., 1933; H. E. McKinstry, *Mining Geology*, 1948.

Ore deposits, geochemistry of

Geochemistry in general deals with the amounts and distribution of the elements and isotopes of the earth, and the nature of the processes affecting them. Although ore bodies are formed by many different processes (see ORE AND MINERAL DEPOSITS), only those deposits of the heavy metals believed to be formed by precipitation from heated water-rich fluids (hydrothermal) are discussed in detail here. Most of the world's supply of base metals, silver, and gold originates in such deposits.

Mineral and chemical composition. The minerals in ore deposits frequently are divided into two groups, ore and gangue, the former constitute those for which the deposit is mined, the latter are the waste minerals associated with the ore. The same mineral may be an ore in some deposits and a gangue in others. The common minerals of hydrothermal deposits (Table 1) are sulfides, sulfo salts, oxides, carbonates, silicates and native elements, although sulfates, a fluoride, tungstates, arsenides, tellurides, selenides, and others are by no means rare. Many minor elements which seldom occur in sufficient abundance to form discrete minerals of their own may substitute for the major elements of the ore minerals and thus be recovered as by-products. For example (as shown in Table 1), the ore mineral of cadmium, indium, and gallium is sphalerite; the major ore mineral of silver and thallium is galena, and pyrite is sometimes an ore of cobalt. See ELEMENTS (GEOCHEMICAL DISTRIBUTION).

Ore deposits consist, in essence, of exceptional concentrations of given elements over that commonly occurring in rocks. The degree of concentration needed to constitute ore varies widely, as shown in Table 2, and is a complex function of many economic and sometimes political variables. The quantity of these elements in the total known or reasonably expected ore bodies in the world is infinitesimal when compared with the total amounts in the crust of the earth. Thus, each and every cubic mile of ordinary rocks in the crust of the earth contains enough of each ore element to make large bodies (see Table 2). Although there is a large number of geologic situations that are apparently favorable, only a very few of them contain significant amounts of ore. Thus it is evident that the processes leading to concentration must be the exception and not the rule, and obviously any understanding or knowledge of these processes should aid in the discovery of further deposits.

It is apparent from the above that each step in the process of ore formation must be examined carefully if this sporadic occurrence of ore is to be placed on a rational basis. In order for ores to form, there must be a source for the metal, a medium in which it may be transported, a driving force to move this medium, a "plumbing system" through which it may move, and a cause of precipitation of the ore elements as an ore body. These interrelated requirements are discussed below in terms of the origin of the hydrothermal fluid, its chemical properties, and the mechanisms by which it may carry and deposit ore elements.

Source of metals. It is not easy to determine the source for the metals in hydrothermal ore deposits because, as shown above, they exist everywhere in such quantities that even highly inefficient processes could be adequate to extract enough material to form large deposits.

Fluids associated with igneous intrusion. In many deposits there is evidence that ore formation was related to the intrusion of igneous rocks nearby, but in many other deposits intensive search has failed to reveal any such association. Because the crystal structures of the bulk of the minerals (mostly silicates) crystallizing in igneous rocks are such that the common ore elements such as copper, lead, and zinc do not fit readily, these elements are concentrated into the residual liquids, along with H_2O, CO_2, H_2S, and other substances. These hot, water-rich fluids, remaining after the bulk of the magma has crystallized, are the hydrothermal fluids which move outward and upward to areas of lower pressure in the surrounding rocks

Fluids obtained from diagenetic and metamorphic processes. Fluids of composition similar to the above also could be obtained from diagenetic and metamorphic processes. When porous, water-saturated sediments containing the usual amounts of hydrous and carbonate minerals are transformed into essentially nonhydrous, nonporous metamorphic rocks, great quantities of water and carbon dioxide must be driven off. Thus, each cubic mile of average shale must lose about 3×10^9 tons of water and may lose large amounts of carbon dioxide on metamorphism to gneiss. The great bulk of the water presumably comes off as connate water (entrapped at time of rock deposition) under conditions of fairly low temperature. In many respects, this water has the same sea-water composition as it had to start with. However, as metamorphism proceeds, accompanied by slow thermal buildup from heat flow from the earth's interior and from radioactivity, the last fluids are given off at higher temperatures and are richer in CO_2 and other substances. These fluids would have considerably greater solvent power and can be expected to

be similar to those coming from cooling igneous rocks.

Role of surface and other circulating waters. It is very likely that the existence of a mass of hot rock under the surface would result in heating and circulation of meteoric water (from rain and snow) and connate water. The possible role of these moving waters in dissolving ore elements from the porous sedimentary country rocks through which they may pass laterally and in later depositing them as ore bodies has been much discussed. The waters may actually contribute ore or gangue minerals in some deposits. The test of this theory of lateral secretion on the basis of precise analyses of the average country rocks around an ore body would involve an exceedingly difficult sampling job. It also would require analytical precision far better than is now feasible for most elements as each part per million uncertainty in the concentration of an element in a cubic mile of rock represents about 11,000 tons of the element or 1,000,000 tons of 1% ore.

Movement of ore-forming fluids. In addition to the high vapor pressures of volatile-rich fluids and

Table 1. Some common primary minerals of hydrothermal ore deposits

Element	Common minerals	Idealized formulas	Significant minor elements occurring in these minerals, underlined where economically important
Iron	Hematite	Fe_2O_3	
	Magnetite	Fe_3O_4	Mn
	Pyrite	FeS_2	Au,* Co, Ni
	Pyrrhotite	$Fe_{1-x}S$	Ni, Co
	Siderite	$FeCO_3$	Mn, Ca, Mg
	Arsenopyrite	$FeAsS$	Sb, Co, Ni
Copper	Chalcopyrite	$CuFeS_2$	Ag, Mn, Se
	Bornite	Cu_5FeS_4	
	Chalcocite	Cu_2S	Ag
	Enargite	Cu_3AsS_4	Ag, Sb
	Tetrahedrite	$Cu_{12}Sb_4S_{13}$	Ag, Fe, Zn, Hg, As
Zinc	Sphalerite	ZnS	Fe, Mn, Cd, Cu, Ga, Ge, Sn, In
Lead	Galena	PbS	Ag, Bi, As, Tl, Sn, Se, Sb
Bismuth	Native bismuth	Bi	
	Bismuthinite	Bi_2S_3	
Silver	Native silver	Ag	Au
	Argentite	Ag_2S	
	Various sulfo salts		
Gold	Native gold	Au	Ag, Cu
	Various tellurides of gold and silver		
Mercury	Cinnabar	HgS	
Tin	Cassiterite	SnO_2	
Uranium	Uraninite	UO_2	Ra, Th, Pb
Cobalt	Cobaltite	$CoAsS$	
	Smaltite	$CoAs_2$	
Nickel	Pentlandite	$(Fe,Ni)_9S_8$	
Tungsten	Scheelite	$CaWO_4$	Mo
	Wolframite	$(Fe,Mn)WO_4$	Mo
Molybdenum	Molybdenite	MoS_2	Re
Manganese	Rhodochrosite	$MnCO_3$	Fe, Mg, Ca
	Rhodonite	$MnSiO_3$	Ca
Others	Calcite	$CaCO_3$	Mn
	Dolomite (and ankerite)	$CaCO_3 \cdot MgCO_3$	Fe, Mn
	Barite	$BaSO_4$	
	Fluorite	CaF_2	
	Quartz	SiO_2	
	Sericite, chlorite, feldspars, clays, and various other silicates		

* Intimately associated as minute particles of metallic gold, but not in the cryst … of the P

Table 2 Approximate concentration of ore elements in earth's crust and in ores

Element	Approximate concentration in average igneous rocks, %	Tons per cubic mile of rock	Approximate concentration in ores, %	Concentration factor to make ore
Fe	5.0	560,000,000	50	10
Cu	0.007	790,000	0.5–5	70–700
Zn	0.013	1,500,000	1.3–13	100–1000
Pb	0.0016	180,000	1.6–16	1000–10,000
Sn	0.004	450,000	0.01*–1	2.5–250
Ag	0.00001	1,100	0.05	5000
Au	0.0000005	56	0.0000015*–0.01	3–2000
U	0.0002	22,000	0.2	1000
W	0.003	340,000	0.5	170
Mo	0.001	110,000	0.6	600

* Placer deposits

ing as a driving force to push them out into the surrounding country rocks and to the surface, there may well be additional pressures from orogenic or mountain-building forces. When a silicate magma has an appreciable percentage of liquid and is subjected to orogenic forces, it moves en masse to areas of lower pressure (it is intruded into other rocks). But if the magma has crystallized 99% or more of its bulk as solid crystals and has only a very small amount of water-rich fluid present as thin films between the grains, and then is squeezed, this fluid may be the only part sufficiently mobile to move toward regions of lower pressure. (If the residual fluid containing the ore elements stays in the rock, it reacts with the early-formed, largely anhydrous minerals of the rock to form new hydrated ones such as sericite, epidote, amphibole [illegible]

[illegible]

[illegible] comprised of joints, faults, porous volcanic plugs, or other avenues. As the fluid leaves the source it moves some appreciable but generally unknown distance laterally, vertically, or both, and finally reaches the site of deposition. This system of channels is of utmost importance in the process of ore formation.

Localization of mineral deposits. It is stated frequently that ore deposits are geologic accidents, yet there are reasons, however abstruse, for the localization of a mineral deposit in a particular spot. One rea[illegible] to the [illegible] adjacen[illegible] major f[illegible] ing for solutions [illegible] may provide plumbing for solutions ascending from unknown depths. Zones of shattering are favored locales for mineralization since these provide plumbing and offer the best possibility for the ore solution to react with wall rock, mix with other waters, and expand and cool, all of which may promote precipitation. Some types of rock, particularly limestone and dolomite, are especially susceptible to replacement and thus often are mineralized preferentially. The chemical or physical properties which cause a rock to be favored by the replacing solutions often are extremely subtle and certainly not understood fully at this time.

Zoning and paragenesis. Mineral deposits frequently show evidence of systematic spatial and temporal changes in metal content and mineralogy that are sufficiently consistent from deposit to deposit to warrant special mention under the terms zoning and paragenesis. Zoning may be on any scale, though the range is commonly on the order of a few hundred to a few thousand feet, and may have either lateral or vertical development. In mining districts such as Butte, Montana, or Cornwall, England, where zoning is unusually well developed, there is a peripheral zone of manganese minerals grading inward through successive, overlapping silver-lead, zinc, and copper zones (and in the case of Cornwall, tungsten, and finally tin). The same sequence of zones appears in many deposits localized about intrusive rocks, suggesting strongly that the tin and tungsten are deposited first from the outward moving hydrothermal solutions and that the copper, zinc, lead, and silver were deposited successively as the solutions expanded and cooled. In other districts, the occurrences of mercury and antimony deposits suggest that their zonal position may be peripheral to that of silver or manganese. The paragenesis, or the sequence of deposition of minerals at a single place, as interpreted from the textural relations of the minerals, follows the same general pattern as the zoning, with the tin and tungsten early and the lead and silver late. With both zoning and paragenesis there are sometimes reversals in the relative position of adjacent zones, and these are usually explained as successive generations of mineralization. Some metals such as iron, arsenic, and gold tend to be distributed through all of the zones, whereas others, such as antimony, tend to be restricted to a single position.

The sequence of sulfide minerals observed in zoning and paragenesis matches in detail the relative abilities of the heavy metals to form complex ions in solution. This observation strongly supports the hypothesis developed later that most ore

transport occurs through the mechanism of complex ions, since no other geologically feasible property of the ore metals or minerals can explain the zoning relations.

Environment of ore deposition. Important aspects of the environment of ore deposition include the temperature, pressure, nature, and composition of the fluid from which ores were precipitated.

Temperatures. Although there is no geological thermometer that is completely unambiguous as to the temperatures of deposition of ores, there is a surprising number of different methods for estimating the temperatures that prevailed during events long since past that have been applied to ores with reasonably consistent results (*see* GEOLOGIC THERMOMETRY). Those ore deposits which had long been considered to have formed at high temperatures give evidence of formation in the range of 500–600°C, or possibly even higher. Those that were thought to be low-temperature deposits show temperatures of formation in the vicinity of 100°C or even less, and the bulk of the deposits lie between these extremes

Pressures. It would be useful to know the total hydrostatic pressure of the fluids during ore formation. Most of the phenomena used for determination of the temperatures of ore deposition are also pressure-dependent, and so either an estimate of the correction for pressure must be made, or two independent methods must be used to solve for the two variables.

Pressures vary widely from nearly atmospheric in hot springs to several thousand atmospheres in deposits formed at great depth. Maximum reasonable pressures are considered to be on the order of that provided by the overlying rock; conversely, the minimum reasonable pressures are considered to be about equal to that of a column of fluid open to the surface. Pressures therefore range from approximately 500 to 1500 psi per 1000 ft of depth at the time of mineralization. *See* HIGH-PRESSURE PHENOMENA.

Evidence of composition. Geologists generally concede that most ore-forming fluids are essentially hot water or dense supercritical steam in which are dissolved various substances including the ore elements. There are three lines of evidence bearing on the composition of this fluid These are fluid inclusions in minerals, thermal springs and fumaroles, and the mineral assemblage of the deposit and its associated alteration haloes

1. Fluid inclusions in minerals. Very small amounts of fluid are trapped in minute fluid-filled inclusions during the growth of many ore and gangue minerals in veins, and these inclusions have been studied intensively for evidence of temperature and composition (F. G. Smith, 1953). Although the relative amounts may vary widely, these fluids will have 5–25 or even more weight per cent soluble salts such as chlorides of Na, K, and Ca, plus highly variable amounts of carbonate, sulfate, and other anions. Some show liquid CO_2 or hydrocarbons as separate phases in addition to the aqueous solution. A few show detectable amounts of H_2S and minor amounts of many other substances. After losing some CO_2 and H_2S through release of pressure and oxidation when the inclusions are opened, the solutions are within 2 or 3 pH units of neutral. There is no evidence of sizable quantities (>1 g/liter) of the ore metals in these solutions, and the evidence indicates that the concentrations of the ore elements must be very low (<0.1 g/liter). Even if the concentrations were in the range of 0.1 g/liter, there should be analytical evidence in the fluid inclusion studies, but this is lacking. In addition, if fluids of such compositions were trapped in fluid inclusions in transparent minerals and on cooling precipitated even a fraction of their metal content as opaque sulfides, these should be visible (under the microscope) within the inclusions, but none are seen. If the concentrations of ore elements are much less than 0.001 g/liter, the volume of fluids that must be moved through a vein to form an ore body becomes geologically improbable.

2. Thermal springs and fumaroles These provide the closest approach to a direct look at the processes of ore deposition as some ore and gangue minerals form within the range of direct observation. The solutions from these springs give diluted and possibly contaminated, partly oxidized and partly devolatilized samples of the sort of fluid that presumably forms ore bodies at greater depths Isotopic studies, for example, using natural tritium as a tracer, and other less quantitative evidence show that the solutions have been diluted by local meteoric water until less than 5–10% of the fluid emitted at the surface is of deep-seated origin. The compositions of these thermal springs, after correction for such dilution, are in good agreement with the data from fluid inclusions *See* RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS.

3. Mineral assemblage. The assemblage of minerals that occurs within a deposit provides a great deal of information about the chemical nature of the fluid from which the ores were precipitated. There are a great number of stable inorganic compounds of the heavy metals known, yet unaltered ore deposits contain only a relatively small number of minerals. For example, lead fluoride, lead chloride, lead carbonate, lead sulfate, lead oxide, lead sulfide, and many others are known stable compounds of lead, yet of these, primary ore deposits contain only the sulfide (galena). Some elements, such as calcium, which occur in combination with several types of anions, for example the carbonate, fluoride, sulfate, and numerous silicates, are found with the ore minerals A quantitative approach to the compositional problem may be made by considering reactions such as

$$CaCO_3 + 2F^- \rightleftharpoons CaF_2 + CO_3^{--}$$

The equilibrium constant for this reaction is $(CO_3^{--})/(F^-)^2 = 10^{1.4}$ at 25°C. Thus when calcite and fluorite are in equilibrium, the requirements for the constant are met,



$(CO_3^{--})/(F^-$

ratio is known. A large number of such equations be evaluated and from comparison with the neral assemblage known to occur in ores, lim- on the possible variation of the composition of e ore-forming fluid may be estimated. Unfortu- tely, calculations of this sort involving ionic uilibria are limited to fairly low temperatures ess than 100-200°C) since there are few relia- e thermodynamic data on ionic species at high mperature. At any temperature, reactions such

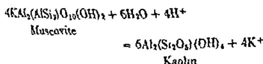


$$2Ag + \tfrac{1}{2}S_2 = Ag_2S$$

an be used to evaluate or place limits on the ossible variation of the chemical potential of ome components in the ore-forming fluid. *See* SUL- IDE PHASE EQUILIBRIA.

The composition of the ore fluid tends to become adjusted chemically by interaction with the rocks with which it comes ...

by such reactions as



$$4KAl_2(AlSi_3)O_{10}(OH)_2 + 6H_2O + 4H^+ \;(\text{Muscovite}) = 6Al_2(Si_2O_5)(OH)_4 + 4K^+ \;(\text{Kaolin})$$

where the equilibrium constant has the form

$$K = \frac{(K^+)^4}{(H_2O)^6\,(H^+)^4}$$

Likewise, the quantitatively small but neverthe- less important partial pressures of sulfur and oxy- gen may be governed by reactions such as



$$Fe_3O_4 + S_2 = Fe_2O_3 + FeS_2 + \tfrac{1}{2}O_2$$

Such changes in the wall rock come under the gen- eral heading of wall-rock alteration and may be of many types, only a few of the more common of which are mentioned below.

High-temperature alteration of limestones usu- ally results in the formation of water-poor cal- cium silicates such as garnet, pyroxenes, idocrase, and tremolite, and the resulting rock is termed skarn. At lower temperatures in the same types of rock, dolomitization and silicification are the pre- dominant forms of alteration, because the partial pressure of CO_2 is too high to permit calcium sili- cate to form. *See* SILICATE PHASE EQUILIBRIA.

At high temperatures in igneous and metamor- phic rocks near granite in composition, the solu- tions are approximately in equilibrium with the primary rock-forming minerals, and thus there is little alteration except development of sericite and occasionally topaz and tourmaline. At lower tem- peratures, the characteristic sequence of altera- tion from fresh rock toward the vein is first an argillic zone, then a sericitic zone, and finally a silicified zone bordering the vein.

Summary. Summarizing the environment of ore deposition, there are various lines of evidence to show that most hydrothermal ore deposits were formed at temperatures of 100-600°C and at pres- sures ranging from nearly atmospheric to several thousand atmospheres. The solutions were domi- nantly aqueous and were fairly concentrated in sodium and potassium chlorides but were rela- tively dilute in terms of the ore metals.

Mechanisms of ore transport and deposition.

The ore minerals, principally the sulfides, are ex- tremely insoluble in pure water at high tempera- tures as well as low; the solubility products are so low, in fact, that literally oceans of water would be required to transport the metal for even a small ore body. Thus, it is not easy to explain the mech- anism whereby the minerals are solubilized to the extent necessary ...

extremely low, the relative solubilities of the sul- fides are radically different. For example, accord- ing to the solubility products, FeS is many, many times more soluble than PbS (about 10^{10} times at 25°C), yet the two minerals occur together in ore deposits and behave as if galena were slightly more soluble than pyrrhotite. From this and other lines of evidence, it appears necessary to conclude that the solubilities of the various contempora- neous minerals in a given deposit could not have differed among themselves by more than a few or- ders of magnitude.

The only geologically and chemically feasible mechanism by which these solubilities may be equalized approximately is the formation of com- plex ions of the heavy metals. Such complexes can increase the solubilities of heavy metals tremen- dously. As an example, the activity (thermody- namic concentration) of Hg^{++} in a solution satu- rated with HgS (cinnabar) and H_2S at 25°C, 1 atm pressure, and pH 8, is only about 10^{-47} moles/ liter, representing a concentration much less than 1 atom of mercury in a volume of water equal to the entire volume of the oceans of the world. However, in the same solution is formed a very stable sulfide complex of mercury, HgS_2^{--}, which increases the total concentration of mercury in solution by the impressive factor of about 10^{42}, giving a concentra- tion on the order of 0.001 g/liter. Not only does complex formation provide a means to achieve adequate solubility for ore transport, but the rela- tive tendency for metals to form certain types of complexes matches in detail the commonly ob- served zoning and paragenetic sequences mentioned previously. The metals whose sulfides are the least soluble tend to form the most stable ...

... geologic importance; a few of the more sig- nificant are sulfide (S^{--}), hydrosulfide (HS^-), polysulfides (S_x^{--}), thiosulfate ($S_2O_3^{--}$), sulfate (SO_4^{--}), carbonate (CO_3^{--}), and chloride (Cl^-).

The precipitation of minerals from complexed solutions takes place either by shifts in equilib-

transport occurs through the mechanism of complex ions, since no other geologically feasible property of the ore metals or minerals can explain the zoning relations.

Environment of ore deposition. Important aspects of the environment of ore deposition include the temperature, pressure, nature, and composition of the fluid from which ores were precipitated.

Temperatures. Although there is no geological thermometer that is completely unambiguous as to the temperatures of deposition of ores, there is a surprising number of different methods for estimating the temperatures that prevailed during events long since past that have been applied to ores with reasonably consistent results (*see* GEOLOGIC THERMOMETRY). Those ore deposits which had long been considered to have formed at high temperatures give evidence of formation in the range of 500–600°C, or possibly even higher. Those that were thought to be low-temperature deposits show temperatures of formation in the vicinity of 100°C or even less, and the bulk of the deposits lie between these extremes.

Pressures. It would be useful to know the total hydrostatic pressure of the fluids during ore formation. Most of the phenomena used for determination of the temperatures of ore deposition are also pressure-dependent, and so either an estimate of the correction for pressure must be made, or two independent methods must be used to solve for the two variables.

Pressures vary widely from nearly atmospheric in hot springs to several thousand atmospheres in deposits formed at great depth. Maximum reasonable pressures are considered to be on the order of that provided by the overlying rock; conversely, the minimum reasonable pressures are considered to be about equal to that of a column of fluid open to the surface. Pressures therefore range from approximately 500 to 1500 psi per 1000 ft of depth at the time of mineralization. *See* HIGH-PRESSURE PHENOMENA.

Evidence of composition. Geologists generally concede that most ore-forming fluids are essentially hot water or dense supercritical steam in which are dissolved various substances including the ore elements. There are three lines of evidence bearing on the composition of this fluid. These are fluid inclusions in minerals, thermal springs and fumaroles, and the mineral assemblage of the deposit and its associated alteration haloes.

1. Fluid inclusions in minerals. Very small amounts of fluid are trapped in minute fluid-filled inclusions during the growth of many ore and gangue minerals in veins, and these inclusions have been studied intensively for evidence of temperature and composition (F. G. Smith, 1953). Although the relative amounts may vary widely, these fluids will have 5–25 or even more weight per cent soluble salts such as chlorides of Na, K, and Ca, plus highly variable amounts of carbonate, sulfate, and other anions. Some show liquid CO_2 or hydrocarbons as separate phases in addition to the aqueous solution. A few show detectable amounts of H_2S and minor amounts of many other substances. After losing some CO_2 and H_2S through release of pressure and oxidation when the inclusions are opened, the solutions are within 2 or 3 pH units of neutral. There is no evidence of sizable quantities (>1 g/liter) of the ore metals in these solutions, and the evidence indicates that the concentrations of the ore elements must be very low (<0.1 g/liter). Even if the concentrations were in the range of 0.1 g/liter, there should be analytical evidence in the fluid inclusion studies, but this is lacking. In addition, if fluids of such composition were trapped in fluid inclusions in transparent minerals and on cooling precipitated even a fraction of their metal content as opaque sulfides, these should be visible (under the microscope) within the inclusions, but none are seen. If the concentrations of ore elements are much less than 0.001 g/liter, the volume of fluids that must be moved through a vein to form an ore body becomes geologically improbable.

2. Thermal springs and fumaroles. These provide the closest approach to a direct look at the processes of ore deposition as some ore and gangue minerals form within the range of direct observation. The solutions from these springs give diluted and possibly contaminated, partly oxidized and partly devolatilized samples of the sort of fluid that presumably forms ore bodies at greater depths. Isotopic studies, for example, using natural tritium as a tracer, and other less quantitative evidence show that the solutions have been diluted by local meteoric water until less than 5–10% of the fluid emitted at the surface is of deep-seated origin. The compositions of these thermal springs, after correction for such dilution, are in good agreement with the data from fluid inclusions. *See* RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS.

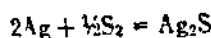
3. Mineral assemblage. The assemblage of minerals that occurs within a deposit provides a great deal of information about the chemical nature of the fluid from which the ores were precipitated. There are a great number of stable inorganic compounds of the heavy metals known, yet unaltered ore deposits contain only a relatively small number of minerals. For example, lead fluoride, lead chloride, lead carbonate, lead sulfate, lead oxide, lead sulfide, and many others are known stable compounds of lead, yet of these, primary ore deposits contain only the sulfide (galena). Some elements, such as calcium, which occur in combination with several types of anions, for example, the carbonate, fluoride, sulfate, and numerous silicates, are found with the ore minerals. A quantitative approach to the compositional problem may be made by considering reactions such as

$$CaCO_3 + 2F^- \rightleftharpoons CaF_2 + CO_3^{--}$$

The equilibrium constant for this reaction is $(CO_3^{--})/(F^-)^2 = 10^{1.4}$ at 25°C. Thus when calcite and fluorite are in equilibrium, the requirements for the constant are met. $(CO_3^{--})/(F^-)^2$



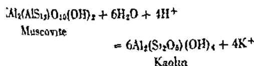
is known A large number of such equations be evaluated and from comparison with the ral assemblage known to occur in ores, lim- the possible variation of the composition of re-forming fluid may be estimated. Unfortu- y, calculations of this sort involving ionic ibria are limited to fairly low temperatures than 100–200°C) since there are few relia- thermodynamic data on ionic species at high perature At any temperature, reactions such



$$2Ag + \tfrac{1}{2}S_2 = Ag_2S$$

be used to evaluate or place limits on the ible variation of the chemical potential of e components in the ore-forming fluid. See SUL- E PHASE EQUILIBRIA.

The composition of the ore fluid tends to become usted chemically by interaction with the rocks h which it comes in contact, and these changes y well contribute to the precipitation of the ore metals. Thus, the K^+/H^+ ratio may be controlled such reactions as



$$Al_2(AlSi_3)O_{10}(OH)_2 + 6H_2O + 4H^+$$

Muscovite

$$= 6Al_2(Si_2O_5)(OH)_4 + 4K^+$$

Kaolin

here the equilibrium constant has the form

$$K = \frac{(K^+)^4}{(H_2O)^6\,(H^+)^4}$$

Likewise, the quantitatively small but nevertheless important partial pressures of sulfur and oxygen may be governed by reactions such as



$$Fe_3O_4 + S_2 = Fe_2O_3 + FeS_2 + \tfrac{1}{2}O_2$$

Such changes in the wall rock come under the general heading of wall-rock alteration and may be of many types, only a few of the more common of which are mentioned below.

High temperature alteration of limestones usually results in the formation of water-poor calcium silicates such as garnet pyroxenes, idocrase, and tremolite, and the resulting rock is termed skarn At lower temperatures in the same types of rock, dolomitization and silicification are the predominant forms of alteration, because the partial pressure of CO_2 is too high to permit calcium silicate to form. See SILICATE PHASE EQUILIBRIA

At high temperatures in igneous and metamorphic rocks near granite in composition, the solutions are approximately in equilibrium with the primary rock-forming minerals, and thus there is little alteration except development of sericite and occasionally topaz and tourmaline At lower temperatures, the characteristic sequence of alteration from fresh rock toward the vein is first an argillic zone, then a sericitic zone, and finally a silicified zone bordering the vein

Summary Summarizing the environment of ore deposition, there are various lines of evidence to show that most hydrothermal ore deposits were formed at temperatures of 100–600°C and at pressures ranging from nearly atmospheric to several thousand atmospheres. The solutions were dominantly aqueous and were fairly concentrated in sodium and potassium chlorides but were relatively dilute in terms of the ore metals.

Mechanisms of ore transport and deposition. The ore minerals, principally the sulfides, are extremely insoluble in pure water at high temperatures as well as low; the solubility products are so low, in fact, that literally oceans of water would be required to transport the metal for even a small ore body Thus, it is not easy to explain the mechanism whereby the minerals are solubilized to the extent necessary for ore transport

In addition to the fact that the absolute solubilities, calculated from the solubility products, are extremely low, the relative solubilities of the sulfides are radically different. For example, according to the solubility products, FeS is many, many times more soluble than PbS (about 10^{10} times at 25°C), yet the two minerals occur together in ore deposits and behave as if galena were slightly more soluble than pyrrhotite. From this and other lines of evidence, it appears necessary to conclude that the solubilities of the various contemporaneous minerals in a given deposit could not have differed among themselves by more than a few orders of magnitude

The only geologically and chemically feasible mechanism by which these solubilities may be equalized approximately is the formation of complex ions of the heavy metals. Such complexes can increase the solubilities of heavy metals tremendously As an example, the activity (thermodynamic concentration) of Hg^{++} in a solution saturated with HgS (cinnabar) and H_2S at 25°C, 1 atm pressure, and pH 8, is only about 10^{-47} moles/liter, representing a concentration much less than 1 atom of mercury in a volume of water equal to the entire volume of the oceans of the world. However, in the same solution is formed a very stable sulfide complex of mercury, HgS_2^{--}, which increases the total concentration of mercury in solution by the impressive factor of about 10^{42}, giving a concentration on the order of 0 001 g/liter. Not only does complex formation provide a means to achieve adequate solubility for ore transport, but the relative tendency for metals to form certain types of complexes matches in detail the commonly observed zoning and paragenetic sequences mentioned previously. The metals whose sulfides are the least soluble tend to form the most stable complexes, and metals whose minerals are comparatively soluble form weaker complexes. There are many kinds of complexing ions or molecules (ligands) of possible geologic importance; a few of the more significant are sulfide (S^{--}), hydrosulfide (HS^-), polysulfides (S_x^{--}), thiosulfate ($S_2O_3^{--}$), sulfate (SO_4^{--}), carbonate (CO_3^{--}), and chloride (Cl^-)

The precipitation of minerals from complexed solutions takes place either by shifts in equilib-

rium caused by changing (usually cooling) temperature or by a decrease in the concentration of the ligand, thereby reducing the ability of the solution to carry the metals. This latter alternative can take place in several ways, as by reaction with wall rock, by mixing with other solutions, or by formation of a gas phase through loss of pressure.

Oxidation and secondary enrichment. When ore deposits are exposed at the surface, they are placed in an environment quite different from that in which they were formed, and the character of the deposit is changed through the processes of oxidation and weathering. The sulfides give way to oxides, sulfates, carbonates, and other compounds which are more or less soluble and tend to be leached away, leaving a barren gossan of insoluble siliceous iron and manganese oxides. Some minerals, such as cassiterite and native gold, may leach away at a less rapid rate than does the surrounding material; thus they are concentrated as a surficial residuum.

Where the country rock is relatively inert to the acid solutions generated by the oxidizing sulfides, as in the case of quartzites and some hydrothermally altered rocks, copper and especially zinc are leached away readily; lead and silver may be retained temporarily in the oxidized zone as the carbonate or sulfate, and the chloride or native metal, respectively; but eventually these too are dissolved away. The various metallic ions are carried downwards until they reach unoxidized sulfides in the vicinity of the water table where the solutions interact with these sulfides to form a new series of supergene sulfide minerals. Copper sulfide is the least soluble sulfide of the plentiful metals in the solution, and hence the zone of supergene sulfide enrichment is predominantly a copper sulfide zone with occasional rich concentrations of silver. Zinc nearly always remains in solution and is lost in the ground water.

In reactive wall rocks, such as limestones, reaction with the wall rock prevents the solutions from becoming acid enough for large amounts of metal to be removed in solution, the base metals are retained almost in place as carbonates, sulfates, oxides, halides, and there is no appreciable sulfide enrichment.

The behavior of some elements is governed by the availability of other materials. Thus, for example, uranium is readily leached from the oxidized zone in many deposits; however, when the oxidizing solutions contain even very small amounts of potassium vanadate, the extremely insoluble mineral carnotite precipitates and uranium is immobilized.

Highly soluble materials, such as uranium in the absence of chemicals that precipitate it, may be temporarily fixed in the oxidized zone by adsorption on colloidal materials such as freshly precipitated ferric oxides.

Current trends in investigation. In recent years there has been a great increase in the degree to which the experimental methods and principles of physical chemistry have been applied to aid in understanding the processes by which ores have formed, and this approach can be expected to be even more fruitful in the future. Several avenues appear promising and are under active investigation in numerous laboratories. Among these are the following.

1. Phase equilibrium studies of both natural and synthetic ore and gangue minerals.
2. Distribution coefficients for trace elements between coexisting phases, and between various forms on the same crystal.
3. Experimental solubility studies in dominantly aqueous solutions.
4. Studies of the composition and origin of thermal spring waters and fluid inclusions in minerals.
5. Thermodynamic properties of minerals.
6. Isotopic fractionation during transportation and deposition processes.
7. Rate studies on crystal growth and habit, diffusion, reaction, transformation, and similar processes.
8. Crystal structure determinations and crystal chemical studies of ore and gangue minerals.
9. Distribution of elements in the earth's crust and in various rock types.
10. Detailed field studies of the relations between minerals in ore deposits.

For a discussion of sensitive chemical analytical techniques used in the search for ore deposits see GEOCHEMICAL PROSPECTING. For further discussion of chemical principles involved in ore deposition, see GEOLOGIC THERMOMETRY; LEAD ISOTOPES, GEOCHEMISTRY OF; LITHOSPHERE, GEOCHEMISTRY OF; SULFIDE PHASE EQUILIBRIA. [E.W.R.; P.B.B.]

Bibliography: P. Abelson (ed.), *Researches in Geochemistry*, 1959; A. M. Bateman (ed.), *Economic Geology, Fiftieth Anniversary Volume, 1905–1955*, 2 vols., 1955; B. Mason, *Principles of* [illegible] and *T. G* [illegible] *istori* [illegible] 1953.

Ore dressing

Treating of ores to concentrate the valuable constituents (minerals) of the ore into a product (concentrate) of smaller bulk, and simultaneously to collect the worthless material (gangue) into a discardable waste (tailing). The fundamental operations of ore-dressing processes are the breaking apart of the associated constituents of the ore by mechanical means (severance), and the separation of the severed components (beneficiation) into concentrate and tailing using mechanical or physical methods which do not effect substantial chemical changes.

Severance. Comminution is a single- or multistage process whereby ore is reduced from run-of-mine size to that size needed by the beneficiation process. It seeks to produce individual particles which are either wholly mineral or wholly gangue, that is, to produce liberation. Since the mechanical forces producing fracture [illegible] susceptible to detailed control, a class [illegible] taini

mineral and gangue (middling particles) are also produced. The smaller the percentage of middlings the greater the degree of liberation Comminution is divided into crushing (down to 6- to 14-mesh) and grinding (down to micron sizes). Crushing is usually done in three stages: coarse crushing from run-of mine size to 4- to 6-in or coarser, intermediate crushing down to about ½-in., and fine crushing to ¼-in or less. *See* SIZE REDUCTION

Screening is a method of sizing whereby graded products are produced, the individual particles in each grade being of nearly the same size (*see* SCREENING). In beneficiation, screening is practiced for two reasons: as an integral part of the separation process, for example, in jigging; and to produce a feed of such size and size range as is compatible with the applicability of the separation process.

Beneficiation. This step consists of two fundamental operations: the determination that an individual particle is either a mineral or a gangue particle (selection); and the movement of selected particles via different paths (separation) into the concentrate and tailing products When middling particles occur, they will either be selected according to their mineral content and then caused to report as concentrate or tailing, or be separated as a third product (middling). In the latter case, the middling is reground, to achieve further liberation, and the product is fed back into the stream of material being treated

Selection is based upon some physical or chemical property in which the mineral and gangue particles differ in kind or degree or both. Thus in hand picking, the oldest form of beneficiation, color, luster and shape are used to decide whether a lump of ore is predominantly mineral or gangue Use is made of differences in other physical or chemical properties such as specific gravity, magnetic permeability, *inductive charging (electrostatic separation)*, surface chemical properties (*see* FLOTATION), bulk chemical properties (*see* LEACHING), weak planes of fracture (separation by screening), and γ ray emission (automatic sorting of radioactive materials) *See* SEPARATION (MECHANICAL).

Separation is achieved by bringing to bear upon each particle of the mixture a set of forces which is usually the same irrespective of the nature of the particles excepting for the force based upon the discriminating property This force may be present for both mineral and gangue particles but differing in magnitude, or it may be present for one type of particle and absent for the other. As a result of this difference, separation is possible, and the particles are collected as concentrate or tailing

Magnetic separat[illegible]
a ma[illegible]
terac[illegible]
Thus under the action of these two forces, different paths are produced for the magnetic and nonmagnetic particles. *Figure 1* shows a continuously moving endless belt B onto which are introduced the particles to be separated The magnetic field is

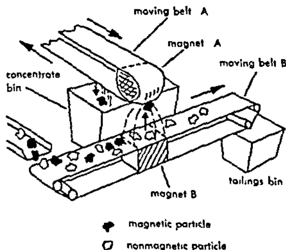


Fig 1. Magnetic separation.

produced by a square-shaped bottom pole B and an upper pole A so curved as to concentrate the lines of force as shown When a magnetic particle comes within the magnetic field, it is attracted strongly enough to move upward against the force of gravity to the surface of an endless belt A conformed to the surface of the pole. The movement of belt A in a direction perpendicular to belt B carries the particle to a concentrate bin. The unattracted nonmagnetic particle being held by gravity continues moving with belt B until it falls into the tailings bin.

All substances placed in a magnetic field acquire magnetic properties. Magnetic permeability is a measure of the ease with which these properties are induced The ratio of permeabilities may be as low as 5 1 for successful separation. However, for such commonly separated materials as magnetite and quartz it is about 110·1

The most important practical separations are those of the *iron ores* In the magnetite ores, magnetite is separated from quartz, feldspars, and the like In the hematite and limonite ores, the ore is first roasted to convert these iron oxides partly into the magnetic oxide and this is then separated from gangue. In the preparation of industrial minerals, magnetic separation is used to clean up, or remove, iron introduced during grinding as in the preparation of china clay, body slip, or glaze, or to remove trace magnetic minerals such as biotite, garnet, and tourmaline from feldspar.

Gravity concentration is based on a discriminating force the magnitude of which varies with specific gravity The other force usually operating in gravity methods is the resistance to relative motion exerted upon the particles by the fluid or semifluid medium in which separation takes place.

Jigging is a gravity method which separates mineral from gangue particles by utilizing an effective difference in settling rate through a periodically dilated bed In Fig 2, the mixture of particles (feed) falls into the jig compartment where it is supported by the screen. Reciprocation of the plunger forces water through the screen and causes

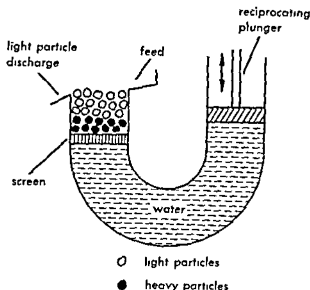


Fig. 2 Jigging

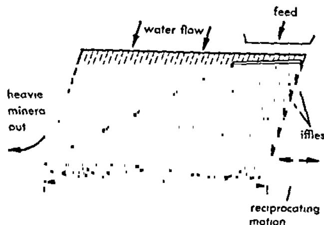
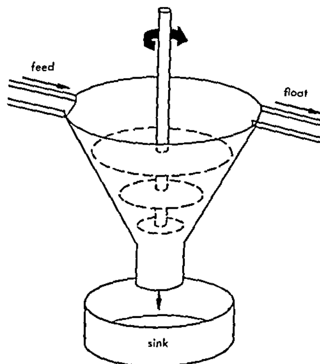


Fig 3 Tabling



periodic dilation and contraction of the bed of particles. During the dilation heavier particles work their way to the bottom while the lighter particles remain on top and are discharged over the lip. Jigging is practiced on materials which are liberated upon being reduced to sizes ranging from 1½ inches down to several millimeters. It has been used on such diverse ores as coal, iron ores, gold and lead ores.

Tabling is a gravity method in which the feed introduced onto a plane, inclined, and reciprocated deck, moves in the direction of motion while simultaneously being washed by a water film which moves it also at right angles to the motion of the deck. In Fig. 3, feed enters at the top of the table and collects within the valleys formed by the narrow cleats, or riffles, which taper in height from right to left. Under the effect of the reciprocating action, the particles stratify with the heavier particles on the bottom and they also move from right to left. Owing to decreasing taper of the riffles, the exposed upper surface of the stratified material is acted upon by cross currents of water and by the tilt of the deck, as indicated by the arrow, and is moved downhill. The heavier mineral and the lighter gangue are usually collected over the edges of the deck as shown. Tables may be used to treat relatively coarse material (sand tables) or fines (slime tables), with sizes ranging from about 2-3 mm down to 0.07 mm.

Sink-float separation is the simplest gravity method based on existing differences in specific gravity. The feed particles are introduced into a suspension the specific gravity of which is between that of the mineral and gangue particles with the result that particles of higher specific gravity sink while those of lower specific gravity float In Fig 4, the separator is a cone equipped with a slowly operated stirrer which serves to impart a slow rotary motion to the suspension and to prevent the suspension from settling out on the walls Feed is introduced at one point of the circumference and is slowly moved by the rotating motion of the suspension By the time this material has reached the discharge point on the circumference, those particles whose specific gravity is greater than that of the suspension have moved down through the suspension so that only float particles are discharged at the top. The sink particles are discharged at the bottom.

The suspension may have a specific gravity ranging from 1 3, using quartz, to 2.4, using galena Magnetite and ferrosilicon are also used at intermediate densities Although there is no top limit to the size of feed particles, a lower limit of about ⅛ in exists with the more standard equipment

Earliest use of sink-float was in the separation of slate from coal using quartz in suspension. A more recent and most important use is to produce at a coarse size a tailing which can be discarded In this manner it is possible to reduce the quantity of material handled by the concentrating plant early in the treatment, th [illegible] ing a

ration is a method of separation based on the ences in size between the things being sepa- Since water is one of the things being sepa- , it has no lower size whereas the solid from it is being separated has a lower size. Conse- tly if a barrier (filter cloth) having openings n can pass water but not the solids is provided, if the pulp is placed on one side of the filter, tion will take place if a pressure is exerted on pulp See FILTRATION, METALLURGY; SOLVENT ACTION. [M.D HA.]

Bibliography: A F. Taggart, *Handbook of Mineral Dressing*, 1945

Organic chemical synthesis

thesis is commonly defined as "composition or putting of two or more things together." As lied to organic chemistry, it refers to the forma- of one product from another, the new sub- ice being usually of higher molecular weight. wever, by usage, it also refers to the prepara- of products of lower molecular weight, fre- ntly through displacement or elimination reac- ns. Conversion of one compound to another linarily requires a reagent and sometimes a alyst; some involve the reaction of two or more lecules of the *same substance*, others the reac- n of two or more molecules of different sub- nces

Applications. Adaptation of chemical reactions the stepwise building of complex compounds of ecise structure from simple compounds is one of e most important facets of organic chemistry om both a theoretical and a practical aspect The ructures of many products isolated from natural urces, previously determined by degradation udies, have been established in this way, for ex- mple, the alkaloids cocaine, morphine, quinine, rychnine, and reserpine, the hormones estrone, estosterone, cortisone, and oxytocin; the antibiotic penicillin V, many flavors and perfumes such as anillin, β phenylethyl alcohol, and dyes such as ndigo and certain anthraquinones Many products can be obtained more economically by synthesis than by extraction from natural sources Consequently, synthetic methods have been adopted for industrial production of such compounds as the dye indigo; the vitamin ascorbic acid, and the flavors oil of wintergreen and oil of orange blossoms

Many synthetically produced organic compounds, which contain combinations of atoms occurring as part of the structures of certain physiologically active natural compounds, have been found to possess the same or higher activity and sometimes lower toxicity than the naturally occurring compounds To illustrate, procaine has largely replaced cocaine as a local anesthetic; Demerol is used frequently in place of morphine as an analgesic.

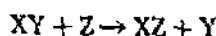
An almost unlimited number of organic compounds of different structures can be made, and many hundreds of thousands have already been described In these are found the synthetic dyes, drugs, fibers, rubbers, and plastics. Eventually, a cure for cancer and other resistant diseases, as well as new articles of commerce superior to those already available, may be discovered among the substances not yet synthesized or among those already known but not yet tested See ORGANIC CHEMISTRY.

Synthetic organic chemistry is the most inclusive branch of this science, for it makes use of the principles developed in all of the others. Especially as it has come to attack more complex problems, has it incorporated more and more of the findings of physical organic chemistry, stereochemistry, and reaction mechanisms

Classification of methods. The methods of conversion of one organic compound to another are numerous. They have been discovered since 1880 and have been developed and established as general reactions through the continuous studies of many investigators. Most of these synthetic methods fall into five general categories, four of which may be expressed in a very simple form A sixth category covers methods belonging in the other categories that are used for synthesis of a large and a special

end product. Although a conversion may appear simple, it is actually complicated in many cases. Several steps, which fall into one or another of the defined categories and which involve unisolated intermediate products, have frequently been demonstrated In spite of the complex mechanisms often encountered, most organic reactions are so well understood that they may be applied readily and successfully

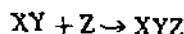
Displacement reactions In these reactions



$$XY + Z \rightarrow XZ + Y$$

one functional group in a compound is replaced by another by a nucleophilic, electrophilic, or free-radical mechanism To illustrate, n-butyl alcohol (n-C_4H_9OH) reacts with phosphorus triiodide (PI_3) to give n-butyl iodide (n-C_4H_9I) and phosphorous acid (H_3PO_3). Bromine (Br_2) and benzene (C_6H_6) react to give bromobenzene (C_6H_5Br) and hydrogen bromide (HBr) 2,4-Dinitrochlorobenzene [2,4-$(NO_2)_2C_6H_3Cl$], an inexpensive and important dye intermediate, reacts with sodium fluoride (NaF) to give 2,4-dinitrofluorobenzene [2,4-$(NO_2)_2C_6H_3F$], which has found utilization in structural organic chemistry. The last reaction represents synthesis of a substance of lower molecular weight See SUBSTITUTION REACTION.

Addition reactions. In these reactions



$$XY + Z \rightarrow XYZ$$

a compound containing an unsaturated functional group combines with a reagent to give a saturated compound When two organic molecules combine in this way, the reaction is frequently classified as a condensation reaction. Propylene (CH_3-$CH{=}CH_2$) will react with hydrogen bromide (HBr) to give either n propyl bromide (CH_3CH_2

CH_2Br) or isopropyl bromide ($CH_3CHBrCH_3$), depending on the conditions. A ketone such as RCOR′ adds hydrogen cyanide (HCN) to give a cyanohydrin [$RR'C(OH)CN$]. Diethyl malonate [$CH_2(CO_2C_2H_5)_2$] and benzalacetophenone ($C_6H_5CH{=}CHCOC_6H_5$) combine to give the addition product [$C_6H_5CH(CH(CO_2C_2H_5)_2)CH_2COC_6H_5$]. *See* ADDITION REACTION.

Elimination reactions. These are the reverse of addition reactions ($XYZ \rightarrow XY + Z$). Some molecule, usually simple in character, is eliminated from a compound with the production of a second compound. Isobutyl chloride [$(CH_3)_2CHCH_2Cl$], upon treatment with ethanolic alkali, loses hydrogen chloride (HCl) with formation of the unsaturated compound isobutylene [$(CH_3)_2C{=}CH_2$]. *tert*-Butyl alcohol [$(CH_3)_3COH$], when heated with an acidic catalyst, loses water with formation of isobutylene [$(CH_3)_2C{=}CH_2$].

Rearrangement reactions The primary product in many organic reactions ($XYZ \rightarrow YXZ$), either after formation or concomitantly in its formation, may rearrange with or without the loss of some simple molecule. Such reactions furnish additional approaches to certain combinations of atoms which are otherwise difficult to attain. Pinacol [$(CH_3)_2COHCOH(CH_3)_2$], when treated with acidic reagents, loses a molecule of water with formation of pinacolone [$(CH_3)_3CCOCH_3$] Allyl phenyl ether ($C_3H_5OC_6H_5$), upon heating, rearranges to *o*-allylphenol (*o*-$C_3H_5C_6H_4OH$).

Condensation reactions Two or more compounds, usually both organic, react together with or without the elimination of a simple molecule to give a new compound Reactions of this type are wide in scope and include some that are frequently classified otherwise. Many of the condensation reactions take place by a complex mechanism involving several steps. Representative types are described and illustrated below.

1 Functional groups from each of two compounds may be removed and the two residues combined to form a new compound Two molecules of bromobenzene (C_6H_5Br) react with copper to give cupric bromide and biphenyl ($C_6H_5C_6H_5$).

2. The same functional groups in two compounds react with each other to give a product of molecular size equal to the sum of the two reacting compounds. Two molecules of aromatic aldehyde (ArCHO) react in presence of cyanide ion to give a benzoin (ArCHOHCOAr).

3. A functional group in one compound reacts with an active position or a functional group in a second compound to bring the two compounds into combination with or without the elimination of some simple molecule Benzaldehyde (C_6H_5CHO) reacts with acetone (CH_3COCH_3) to give benzalacetone ($C_6H_5CH{=}CHCOCH_3$) and water, through an unstable intermediate hydroxy ketone adduct ($C_6H_5CHOHCH_2COCH_3$). An aliphatic aldehyde (RCHO) reacts with a Grignard reagent (R′MgBr) to form an adduct [RCH(OMgBr)R′] which hydrolyzes to an alcohol (RCHOHR′). *See* CONDENSATION REACTION.

Polymerization reactions. Many molecules of one or more simple compounds react to form giant molecules or polymers, with or without the elimination of some simple substance. Two general types of reactions provide the means of synthesis of the vast majority of polymers. These are illustrated below

1. A simple compound, usually in the presence of an initiator, may give a homopolymer Ethylene ($CH_2{=}CH_2$), in the presence of an appropriate initiator, is converted to a plastic polymer, polyethylene ($—CH_2CH_2—)_x$. Two different simple compounds with the same functional group may react to give a copolymer with either regularly recurring or randomly distributed units Vinyl chloride ($CH_2{=}CHCl$) polymerizes with vinylidene chloride ($CH_2{=}CCl_2$) to make a valuable product, Saran [$(—CH_2—CHCl)_x—(CH_2—CCl_2—)_y$] Such polymerization reactions involve a succession of reactions of the simple molecule or molecules with each other.

2. Compounds each of which have two or more functional groups, may react, with or without the elimination of some simple molecule, to give a condensation polymer. Hexamethylene diamine [$H_2N(CH_2)_6NH_2$] and adipic acid [$HO_2C(CH_2)_4CO_2H$] react in equimolecular quantities to give water and a polyamide [$—HN(CH_2)_6NHCO(CH_2)_4CO—]_x$, commonly known as nylon.

See POLYMERIZATION; *see also* ORGANIC REACTION MECHANISM; STEREOCHEMISTRY; STERIC EFFECT (CHEMICAL REACTIONS). [R A]

Bibliography: R. Adams (ed), *Organic Reactions*, vols 1–10, 1941–1959; *Organic Syntheses*, vols 1–38, 1941–1958.

Organic chemistry

The chemistry of the compounds of carbon Intensive development of this branch of chemistry began about the middle of the nineteenth century. However, a lapse of several decades occurred between the beginning of the science and the emergence of a clean-cut definition. Despite the comparatively recent development of organic chemistry as a separate branch of the broader field of chemistry, many typical organic compounds have been known and used for centuries. The Old Testament contains numerous references to the physiological effect of ethyl alcohol, a typical organic compound, as a component of fermented grape juice and to the properties of acetic acid present in what is now called wine vinegar Certain natural dyestuffs, for example, indigo and alizarin, were known to the Egyptians, and the poisonous properties of the hemlock, now known to be ascribable to the alkaloid coniine, were known in the Greek city states. Socrates used an extract of poison hemlock to end his life.

The comparatively late development of organic chemistry is due to the fact that most organic compounds found in nature occur in plant and animal materials as complex mixtures Methods for separation and isolation of the pure compounds have become available only d the past two or three centuries. By the la nineteenth

ntury, an impressive number of organic com-
unds had been isolated from natural sources
nong these are alcohol, urea, uric acid, and many
the organic acids.

Inasmuch as all of these substances were derived
om one or another living organism, the idea de-
loped that some vital force was required for the
nthesis of compounds by such organisms and the
rm organic chemistry was coined to denote that
anch of chemistry which dealt with the products
living organisms. Despite the fact that the Ger-
an chemist Friedrich Wohler succeeded in syn-
esizing urea, a typical organic compound, from
nmonium cyanate without the intervention of a
tal force in 1828, the concept of the necessity for
ich a force in the synthesis of organic compounds
ersisted for several years Eventually this concept
as abandoned and the modern concept of organic
nemistry, embracing the chemistry of the carbon
ompounds, emerged

From the early 1800s, the development of the
cience was rapid. A firm foundation for the subse-
uent rapid advances was furnished by the intro-
uction of quantitative analytical methods applica-
le to the analysis of organic compounds by Justus
iebig and Jean B A Dumas and by the develop-
ent of structural theory in the minds of Stanislao
annizarro and Friedrich A Kekule

Organic chemistry owes its peculiar and impor-
ant position to the fact that *carbon, almost alone
among the elements*, is capable of uniting with it-
elf indefinitely to form compounds Other ele-
nents, notably boron, display similar tendencies,
but carbon forms a far greater number of com-
pounds Secondly, carbon, almost without excep-
tion, displays a constant valence of 4 On these two
principles the science of organic chemistry is built
The number of carbon compounds theoretically
capable of existence is staggering and this very fact
poses problems of major magnitude in connection
with nomenclature, molecular structure, and ar-
rangement in space of the atoms of organic mole-
cules.

Organic compounds in general differ from inor-
ganic compounds by the nature of the bonds by
which the component atoms of a molecule are
united The valence bonds of most inorganic com-
pounds are of the ionic or electrovalent type in
which the outer valence electron shells are filled to
a noble gas arrangement by gain or loss of elec-
trons from the constituent atoms with resultant
development of charged species (ions) Figure 1
shows a schematic electronic representation of so-
dium chloride, a typical inorganic compound

In contrast, when the outer valence electron shell
is filled to a noble gas configuration by sharing
rather than by transfer of electrons between two
atoms the bond so formed is known as a covalent or
electron pair bond These bonds occur in some in-
organic compounds such as ...
tically a

This c
fests it ... striking differences between the
physical and chemical properties of inorganic com

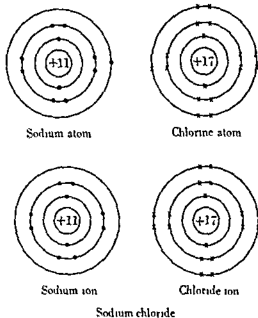


Fig 1. Ionic or electrovalent bond

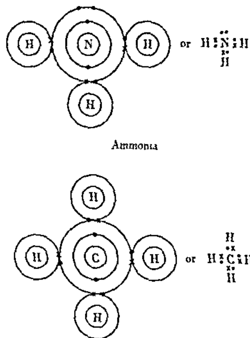
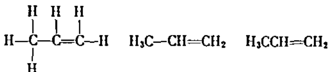


Fig 2 Covalent or electron pair bond.

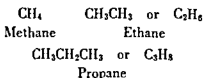
pounds, such as salts, and those of typical organic
compounds. Thus, the inorganic salts in general
have high melting points, cannot be distilled, are
readily soluble in water, but are sparingly soluble
or insoluble in nonaqueous solvents, and conduct
electricity when molten or in aqueous solution. In
contrast, organic compounds possess relatively low
melting points, can frequently be distilled, are
sparingly soluble or insoluble in water but soluble
in nonaqueous solvents, and do *not* conduct elec-
tricity

Over the years, certain conventions for expressing molecular structures of organic compounds have grown up. Actual indication of the shared electrons as in the above examples is cumbersome. Standard practice represents a shared electron bond involving a single pair of electrons (a single bond) by a single line. Double bonds, involving two pairs of shared electrons, are represented by a double line, and triple bonds, involving three pairs of shared electrons, are represented by a triple line. *See* RESONANCE (MOLECULAR STRUCTURE). Further simplification is frequently achieved by omission of bond lines entirely when the meaning is obvious. This progressive simplification is illustrated with the hydrocarbon propene.



An organic compound which contains only single bonds between carbon atoms is said to be saturated If a compound contains one or more multiple bonds between carbon atoms, it is said to be unsaturated

The ability of carbon to combine with itself leads to the existence of series of compounds, the formulas of which differ by a constant increment. Such a series, known as a homologous series, is illustrated by the alkanes, each member of which differs from the next lower by the increment CH_2

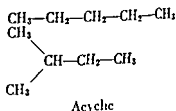


CLASSIFICATION OF ORGANIC COMPOUNDS

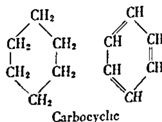
Because of the great number and variety of organic compounds (well over 1,000,000 are known and the number identified is constantly increasing), some systematic classification scheme and systematic method of nomenclature for dealing with them becomes mandatory The problem is complex, and no completely satisfactory system has yet been devised.

In the early days, organic compounds were given names which for the most part were derived from the names of the natural sources of the compounds. This practice has carried down to the present and the use of such trivial names, although convenient, is of little help as far as systematic nomenclature is concerned. Further, at the time of the first isolation of a compound from a natural source, nothing is known about its molecular structure. Adoption of a trivial name is then almost mandatory. However, when the molecular structure of a compound becomes known, assignment of a logical systematic name becomes *possible*, at least in principle. After many attempts to develop a rational system of nomenclature, the matter was finally placed in the hands of a committee of the International Union of Pure and Applied Chemistry. From the efforts of this committee, an international system, the Geneva or IUC system, is gradually emerging, by which organic compounds can be named in a logical manner. However, a certain amount of nationalism still persists and practice does not uniformly conform to the Geneva rules.

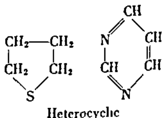
Carbon atoms may combine in the form of long chains, either straight or containing branches, or they may combine to form rings. Further, since carbon is also capable of covalent bonding with other atoms, incorporation of such so-called hetero atoms into carbon rings is possible This situation furnishes the basis for a broad classification of organic compounds into three main groups depending upon the arrangement of the carbon atoms and the presence or absence of atoms other than carbon in the cyclic compounds. Under these terms of reference, organic compounds may be classified as acyclic compounds, which contain no ring structural arrangements of the constituent atoms; carbocyclic compounds, which contain one or more rings consisting solely of carbon atoms; and heterocyclic compounds, the rings of which contain one or more atoms other than carbon. These principles are illustrated by the following examples:



Acyclic



Carbocyclic



Heterocyclic

Obviously little progress results by merely subdividing an unwieldy group of compounds into three almost equally unwieldy groups Further subdivision is mandatory.

Compounds which contain only carbon and hydrogen are known as hydrocarbons. In the hydrocarbons, one or more of the hydrogen atoms, at least in principle, may be replaced by any other atom or group of atoms capable of entering into a covalent bond. Application of this principle furnishes a sound basis for a systematic scheme of further subdivision and nomenclature of the organic compounds. Examples of co[illegible] in which such substitution has been pe[illegible] followin :

$CH_3—CH_3$	Parent compound
H $CH_3—C:Cl$ H	Substitution of hydrogen by chlorine
H $CH_3—C:CH_3$ H	Substitution of hydrogen by $—CH_3$
H $CH_3—C \cdot OH$ H	Substitution of hydrogen by —OH
H $CH_3—C:NH_2$ H	Substitution of hydrogen by $—NH_2$

If the substituent group of atoms is formed from a member of any of the three main classes of organic compounds by loss of one or more atoms of hydrogen, it is known as a radical. An example would be the formation of the methyl radical $—CH_3$, from methane, CH_4. If the substituent is a single atom or group of atoms other than a radical, it is known as a functional group, the presence of which confers characteristic chemical properties on the molecule bearing it. Double or triple carbon-to-carbon bonds also come under the heading of functional groups, since their presence changes the chemical properties of the substituted compound from those of the parent substance (usually the saturated hydrocarbon).

The common functional groups are listed below

Name	*Structure*
Halo (chloro, bromo, etc.)	—Cl, —Br
Hydroxyl	—OH
Aldehyde	H —C=O
Carboxyl	O ‖ —C—OH
Ketone	>C=O
Ether	—O—
Amino	$—NH_2$
Cyano	—C≡N
Thiol or mercapto	—SH
Sulfonic acid	$—SO_3H$

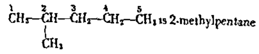
NOMENCLATURE OF ORGANIC COMPOUNDS

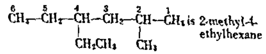
Acyclic hydrocarbons. The acyclic hydrocarbons are subdivided on the basis of the presence or absence of doubl[illegible] containing [illegible] or paraffins [illegible] bonds are [illegible] as alkenes or olefins; and those containing one or more triple bonds are known as alkynes or acetylenes.

Alkanes and alkyl radicals. Alkanes are named by reference to the longest straight chain of carbon atoms present. The names always terminate in "ane." Straight-chain alkanes containing 1, 2, 3, or 4 carbon atoms are known by the trivial names methane, ethane, propane, and butane, respectively. Higher members of the series bear a prefix indicating the number of carbon atoms in the alkane. Thus, $CH_3(CH_2)_nCH_3$ is hexane, dodecane, or tetracontane, when n is 4, 10, or 38, respectively.

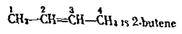
Loss of one hydrogen atom from an alkane results in formation of an alkyl radical which is named by replacement of the suffix "ane" of the parent alkane by "yl." Thus $—CH_3$ derived from methane by loss of one hydrogen is a methyl radical; $CH_3—CH_2—$, similarly derived from ethane, is an ethyl radical. The problem of naming radicals derived from alkanes higher than ethane is complicated, since the hydrogens which may be lost from the alkane are not equivalent. See ISOMERISM, MOLECULAR

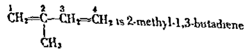
Branched chain alkanes are named on the following principle. The longest continuous straight carbon chain represents the parent alkane and the carbon atoms are numbered consecutively, beginning with one end of the chain. The positions of substituent radicals are then indicated by reference to the number of the carbon atom of the parent alkane to which they are linked. Numbering of the parent alkane is done in such fashion that numbers of the carbon atoms bearing the substituents are the lowest. Thus





Alkenes. Alkenes are named by replacing the terminal "ane" of the parent alkane by "ene." The position of the double bond is indicated by a number indicating from which carbon atom of the longest continuous straight chain the double bond proceeds. The lowest number rule obtains as with branched chain alkanes. Thus





Alkynes. Alkynes are named by replacing the terminal "ane" of the parent alkane by "yne." Otherwise, the rules applying to alkenes hold.

Acyclic hydrocarbon derivatives. In general, the same principles which govern the naming of branched-chain alkanes, alkenes, and alkynes hold. The parent compound is that which contains the longest continuous straight carbon chain. The posi-

tion of functional groups is indicated by numbers, with the smallest number principle again controlling, and the nature of the functional substituent is indicated by suitable prefixes or suffixes.

Halogen compounds. The prefixes fluoro, chloro, bromo, and iodo are used with the name of the parent alkane:

$CH_3CHClCH_2CH_3$ is 2-chlorobutane
$CH_3CHBrCH_2CH_2CH_2Br$ is 1,4-dibromopentane

Hydroxyl derivatives These substances are commonly known as alcohols and many of them are named as such. CH_3OH is methyl alcohol, and CH_3-CH_2OH is ethyl alcohol. In the Geneva system they are named by replacing the terminal "e" of the parent longest straight-chain alkane by "ol" and by indicating the position occupied by the hydroxyl function by a suitable number.

The carbon chain is numbered from the end which gives the carbon bearing the hydroxyl group the smallest number. Still a third scheme names the higher alcohols as derivatives of the first member of the series, carbinol. Thus

CH_3OH is methyl alcohol (carbinol, methanol)

$CH_3CH_2CHCH_2CH_3$ (with OH on the third carbon) is 3-pentanol (diethylcarbinol)

CH_3CHCH_2OH (with CH_3 on the second carbon) is 2-methyl-1-propanol (isopropylcarbinol)

Carbonyl derivatives. The functional group $>C{=}O$ is known as the carbonyl group If one of the unsatisfied valences of carbon is linked to a radical and the other is satisfied by hydrogen, the resulting substance is known as an aldehyde, whereas if the valences of carbon are satisfied by two radicals, the resulting substance is a ketone.

Aldehydes, particularly the lower members of the series, are frequently named by replacing the terminal "ic" of the related acid by "aldehyde." Thus CH_3CHO is acetaldehyde by virtue of its relationship to acetic acid, CH_3COOH. By the Geneva system, aldehydes are named by replacing the terminal "e" of the parent alkane by "al." The aldehyde carbon atom is considered part of the longest straight carbon chain. Since the aldehyde function of necessity must embrace a terminal carbon atom of the parent alkane, it automatically becomes carbon atom 1 of the chain, so that it is unnecessary to specify its position. Thus

$CH_3CHCH_2CH_2CHO$ (with CH_3 on the second carbon) is 4-methylpentanal

Ketones are named by three methods. The lower symmetrical ketones are named by replacing the terminal "ic" of the acid which would yield them on pyrolysis by "one" (see KETONE). Thus acetic acid, CH_3COOH, on pyrolysis yields acetone, CH_3-$COCH_3$. A second scheme, convenient for naming mixed ketones, prefixes the word ketone by the names of the radicals joined to the carbonyl carbon atom. Thus $CH_3COCH_2CH_3$ is methyl ethyl ketone. In the Geneva system the terminal "e" of the parent alkane is replaced by "one." The position of the carbonyl group must obviously be shown by a number. Thus

$CH_3CH_2CCH_2CH_3$ (with $=O$ on the third carbon) is 3-pentanone

Carboxyl functions. The carboxyl function is characteristic of organic acids. Since many acids were originally isolated from natural products, trivial names associated with their sources are common. For example, $HCOOH$, formic acid, is named from the Latin *formica* (ant), since it was first isolated from ants. In general, trivial names are given only to the straight-chain acids. Branched chain acids are named frequently as derivatives of acetic acid, CH_3COOH, or as derivatives of the parent straight-chain acid. Thus

C_2H_5—$CHCOOH$ (with C_2H_5 on the CH) is diethylacetic acid

$CH_3\overset{\beta}{C}H$—$\overset{\alpha}{C}HCOOH$ (with CH_3 on each of the β and α carbons) is α,β-dimethylbutyric acid

By the Geneva system, acids are named by replacing the terminal "e" of the parent hydrocarbon by "oic acid." The carbon atom of the carboxyl group is always numbered 1 and the longest straight chain bearing the carboxyl group controls Thus

$\overset{5}{C}H_3$—$\overset{4}{C}H_2$—$\overset{3}{C}H$—$\overset{2}{C}H_2$—$\overset{1}{C}OOH$ (with OH on carbon 3) is 3-hydroxypentanoic acid

Ethers Ethers are generally named by reference to the alkyl radicals present and addition of the word ether; CH_3CH_2—O—CH_3 is ethyl methyl ether. By the Geneva system they are named as alkoxy derivatives of the parent hydrocarbon. Thus

$CH_3CH_2CHCH_2CH_2CH_3$ (with OCH_2CH_3 on the third carbon) is 3-ethoxyhexane

Amines Amines are alkyl derivatives of ammonia. They are classified into primary, secondary, or tertiary amines according to whether 1, 2, or 3 hydrogen atoms of ammonia have been replaced by organic radicals They are conveniently named by using the names of the radicals attached to nitrogen as prefixes to the word amine Thus

CH_3CH_2—NH—CH_3 is ethylmethylamine

The NH_2 group is known as an amino group and primary amines may therefore be named by treating such a group as a substituent. Thus

$CH_3CHCH_2CH_2CH_2OH$ (with NH_2 on the second carbon) is 4-amino-1-pentanol

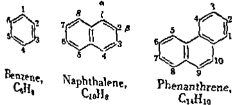
Thiols or mercaptans When the SH group is at the terminus of a chain, it is convenient to prefix the word mercaptan by the radical with which it is joined Thus CH_3CH_2SH is ethyl mercaptan The SH group in the Geneva system is treated as a substituent Thus ethyl mercaptan is ethanethiol.

Sulfonic acids. These are universally named by an additive system The term sulfonic acid is added as a suffix to the name of the parent hydrocarbon regardless of the class to which it belongs. CH_3SO_3H then is methanesulfonic acid

Carbocyclic compounds. The carbocyclic com-

Aromatic compounds. The aromatic carbocyclic compounds are characterized structurally by an alternating system of single and double bonds For the present purpose this somewhat oversimplified picture will suffice (*see* AROMATIC HYDROCARBON). The term aromatic arose because many of these substances were originally isolated from aromatic substances, such as tolu balsam, oil of wintergreen, or oil of cloves

The parent aromatic hydrocarbons are almost always known by trivial names, many of which indicate the source from which they were originally obtained For convenience, it is not customary to write the individual carbon and hydrogen atoms in the structural formulas of aromatic hydrocarbons Rather, each angular position is assumed to represent a carbon atom and the presence of sufficient hydrogen atoms to bring the carbon atoms to the quadrivalent state is also assumed Typical examples are



Benzene, C_6H_6 Naphthalene, $C_{10}H_8$ Phenanthrene, $C_{14}H_{10}$

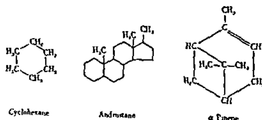
Phenyl, C_6H_5— Naphthyl, $C_{10}H_7$— Phenanthryl, $C_{14}H_9$—

Radicals

Just as acyclic hydrocarbons can give rise to radicals by loss of hydrocarbo one or more of one hydr from benzene is a phenyl radical, and since all hydrogens are equivalent, it is irrelevant which hydrogen is lost With naphthalene and other polynuclear aromatic hydrocarbons, the positions are not equivalent, and it becomes necessary to specify from which position the hydrogen is missing This is conventionally done by use of numbers, or in the case of naphthalene, by the use of Greek letters For example, it is necessary to qualify the names of the two types of naphthyl radicals by use of the terms 1 naphthyl (or α-naphthyl) and 2 naphthyl (or β-naphthyl)

Functional group substituents in aromatic compounds are generally indicated by appropriate adjectives, such as chlorobenzene and nitronaphthalene The positions occupied by substituents are indicated, when necessary, by numbers. Monohydroxy derivatives of benzene and naphthalene are commonly known as phenol and naphthol. Polyhydroxy derivatives of benzene generally bear trivial names (*see* PHENOL).

Alicyclic compounds These are cyclic compounds with properties resembling acyclic rather than aromatic substances. They range in complexity all the way from simple mononuclear, or single-ring, hydrocarbons to the complex multiring substances found in the terpenes and steroids Representative examples are shown in the following structures:



Cyclohexane Androstane α-Pinene

One or more double bonds may be present in an alicyclic compound, provided they do not occur in an aromatic arrangement.

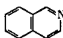
The problem of nomenclature of the alicyclic compounds is formidable and is complicated by the occurrence of structural, positional, geometrical, and stereo isomerism The simple monocyclic hydrocarbons are generally named by use of the prefix "cyclo" in conjunction with the names of the alkane of the same number of carbon atoms. Many alicyclic compounds bear trivial names derived from the natural sources of the substances Definitive rules for naming and numbering many of the polycyclic compounds have been formulated by the International Union of Pure and Applied Chemistry (*see* ALICYCLIC

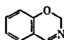
the acyclic series.

Heterocyclic compounds. One or more carbon atoms (with or without accompanying hydrogens) in a carbocyclic compound, may be replaced by a hetero atom (sometimes in combination with hydrogen), provided that the valence requirements are identical The resulting substance is a heterocyclic compound. Examples of such replacements are found in replacement of a —CH= in benzene by —N=, of —CH_2 in cyclopentane by —S—, —O—, or —NH— Although any atom or group of atoms which satisfies the valence requirements may be found in a heterocyclic compound, the most important hetero atoms are O, S, and N

A large number of heterocyclic compounds are commonly known by trivial names, many of which reflect their origin (*see* HETEROCYCLIC COMPOUNDS) A systematic method which can be applied to the nomenclature of any heterocyclic compound has been developed. In this the related aromatic hy-

drocarbons are taken as reference standards, the nature of the hetero atom(s) is indicated by a prefix (oxa- for O, thia- for S, and aza- for N) and the position of the hetero atom or atoms is indicated by a number corresponding to the accepted numbering of the aromatic hydrocarbons. Thus



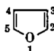


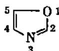
Trivial name. Isoquinoline None

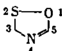
Systematic name 2 Azanaphthalene 1 Oxa-3-aza *2H* naphthalene

In compounds carrying more than one hetero atom, numbers are selected so that oxygen bears the lowest number followed by sulfur and, finally, by nitrogen. The most highly unsaturated compound is taken as the parent Derivatives in which one or more double bonds are saturated are referred to as dihydro or tetrahydro derivatives of the parent.

No uniformity exists in numbering the positions in heterocyclic compounds. In monocyclic substances, numbers are chosen in such fashion that a single hetero atom bears the number one; if more than one hetero atom is present, oxygen bears the number one, followed by sulfur and nitrogen. Thus







With polyheterocyclic compounds the numbering system is confused. European practice frequently differs from American practice, and the American conventions are undergoing constant revision. The safest practice is to write the complete structural formula and make a visual comparison of the positions of substituents. *See* ALCOHOL; ALDEHYDE, AMINE; CARBOHYDRATE; CARBOXYLIC ACID, ETHER; HALOGENATED HYDROCARBON; HYDROCARBON; ORGANOSULFUR COMPOUND; TERPENE [R.C.E.]

Bibliography: J English, Jr , and H. G. Cassidy, *Principles of Organic Chemistry*, 2d ed., 1956; International Union of Pure and Applied Chemistry, Tentative rules for organic nomenclature, *Compt. rend. conf. union intern. chim. pure et appl., 18th conf.*, vol. 160, 1955; F. J. Moore, *A History of Chemistry*, 3d ed , 1939, A M. Patterson and L. T. Capell, *The Ring Index*, 1940.

Organic quantitative analysis

The determination of elements, functional groups, or molecules in organic materials. The type of analysis used is determined by the information required If the total nitrogen content is desired, elemental analysis is used; if only amino nitrogen is desired, then only the amino group is determined. Organic analyses are made on a wide range of materials from pure compounds to mixtures such as [illegible] Selection of the proper method [illegible]

[illegible] may be done in two ways. *In the first method*, the sample is moistened with sulfuric acid and is heated in oxygen to obtain metal oxides or sulfates, which are weighed. Some metals such as gold and silver are weighed as metals. The second method is based on destruction of the organic portion of the sample by heating with nitric and sulfuric acids, followed by determination of the metals by regular procedures. This method is required for metals which are easily volatilized, such as arsenic.

Carbon and hydrogen. These elements are usually determined simultaneously by burning a sample in a stream of oxygen to form carbon dioxide and water. This is done at 600°C in the presence of platinum as a catalyst. If nitrogen is present in the sample, the gas stream is passed through lead peroxide to remove the oxides of nitrogen. Silver wool adsorbs halogens and oxides of sulfur. The water formed is absorbed on a drying agent such as magnesium perchlorate, and from the increase in weight, the amount of hydrogen in the sample can be calculated. Carbon dioxide is absorbed by an alkaline solid such as sodium hydroxide on asbestos fibers, and from the increase in weight, the amount of carbon can be calculated. Carbon alone is determined by wet combustion. The sample is heated in a mixture of sulfuric acid and potassium or silver dichromate. The carbon dioxide formed is determined from the amount of sodium hydroxide with which it combined, or by measuring directly the volume of the carbon dioxide.

Oxygen. Direct determination of oxygen in organic materials is accomplished by heating the sample in a stream of nitrogen to form water, oxides of carbon, and hydrocarbons. On passage of this mixture through graphite at 1150°C, all the oxygen is converted to carbon monoxide. The carbon monoxide is oxidized with iodine pentoxide to form carbon dioxide and iodine Either the iodine or the carbon dioxide may then be measured. This procedure is difficult to perform, so that oxygen in pure compounds is usually obtained by difference rather than by direct measurement

Nitrogen. Two methods are used to determine nitrogen In the Dumas method, the sample is mixed with copper oxide and is heated in a stream of pure carbon dioxide. The elemental nitrogen formed is collected over 50% potassium hydroxide solution, and its volume is measured. This procedure determines total nitrogen in most samples. In the Kjeldahl method, the sample is heated in concentrated sulfuric acid containing a catalyst; this procedure converts the nitrogen to ammonia On the addition of sodium hydroxide followed by boiling, ammonia is distilled into dilute boric acid solution, which is then titrated with acid The choice of the catalyst is critical because different nitrogen compounds require a variety of catalysts. Esters of nitric and nitrous acids cannot be analyzed by this procedure.

Other elements. Sulfur, halogens, and phosphorus are determined by conversion to sulfuric acid, halogen acids, and phosphoric acids, which are measured by standard inorganic procedures In the Carius method, the sample is heated with fuming nitric acid in a sealed t[illegible] the combustion method, the sample is [illegible] en, oxide[illegible]

ir are absorbed in neutral hydrogen peroxide, halogens are absorbed in sodium bisulfite solu- In the bomb method for sulfur, the sample is ed with sodium peroxide and sugar.

inctional groups. Groups such as carboxyl, roxyl, nitro, and amide, are determined by nical reactions which are characteristic for group These reactions include neutralization, dation reduction, precipitation, condensation, gas evolution. These methods are based on sumption of a reagent measured by direct titra- or, in some cases, by determination of excess gent. For others, a product formed in the reac- n is measured Since most organic compounds insoluble in water, organic solvents are used ese solvents enhance the acidity or basicity of ne groups and also permit some reactions which e not possible in water Therefore, the use of non- ueous solvents is important in organic analysis.

Other methods. Many organic compounds ab- rb energy in the ultraviolet or infrared regions nce this absorption is characteristic for each pe of molecule, a direct analysis based on energy bsorption is frequently possible The variation in iass among hydrocarbons is the basis for the mass- pectrometric analysis of petroleum samples These nd other instrumental methods are based on the roperties of molecules

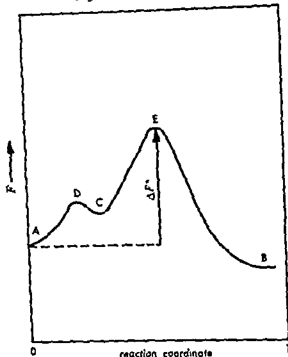
Physical methods of separation are commonly re- quired to obtain molecules free from interferences Normal distillation, vapor distillation, azeotropic distillation, ion exchange, chromatography, extrac- tion of solids and liquids, and diffusion have a place in quantitative analysis. [K.C.S.]

Bibliography. N H. Furman (ed.), 5th ed., *Scotts' Standard Methods of Chemical Analysis* 1939; J Mitchell et al. (eds), *Organic Analysis*, 3 vols., 1953–1956; J Niederl and V. Niederl, *Micro-methods of Quantitative Organic Analysis*, 1942

Organic reaction mechanism

A pathway of chemical states traversed by an or- ganic chemical system in its passage from react- ants to products In the illustration, a plot of free energy vs. reaction coordinate for one kind of re- acting system is shown, here free energy has its usual thermodynamic significance and reaction co- ordinate may be taken to mean the fraction of B-(product) character that an A-(reactant)-mole- cule has acquired The free energy of activation ΔF^* is defined as $\Delta F^* = -RT \ln k$, where k is the reaction rate at unit concentration of the react ants, R is the molar gas constant, and T is the ab- solute temperature See KINETICS (CHEMICAL)

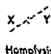
A complete description of the reaction mecha- nism requires not only a knowledge of the struc- tures of the stable molecules A and B but also a knowledge of the structures of any metastable in- termediates, for example, C, and of the structures of the transition states, D and E. Because the transi- tion states D and E, by definition, are very short- lived it is hopeless to attempt to isolate and han- dle them by conventional chemical techniques The state C is also usually short lived However, in-

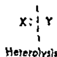


Reaction coordinate diagram.

formation on the structures of C, D, and E can of- ten be inferred from indirect evidence

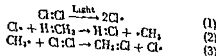
The kinds of evidence that can be used for the elucidation of reaction mechanism include (1) kinetics, that is, the mathematical form of the rate law, the free energy of activation, ΔF^*, and the ef- fect of changes in structure of the reactants or changes in reaction conditions upon the rate of re- action; (2) isolation or trapping of metastable in- termediates; (3) effect of changes in structure of the reactants upon the product distribution in a given type of reaction, (4) the stereochemistry of the process; (5) information obtained by isotopic tracer techniques





Organic reactions usually involve the breaking of covalent bonds between atoms such as X and Y. Usually, the break occurs either homolytically (each atom acquiring one electron of the bonding pair) or heterolytically (one of the atoms acquir- ing both electrons and the other being left with an electron deficiency). There is also a group of re- actions in which the bond breaking processes are less clearly definable

Homolytic reactions. An example of the homo- lytic type is the light induced halogenation of al- kanes. For example, the chlorination of methane may be represented by the steps:



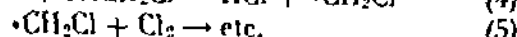
$$Cl{:}Cl \xrightarrow{\text{Light}} 2Cl\cdot \quad (1)$$

$$Cl\cdot + H{:}CH_3 \rightarrow H{:}Cl + \cdot CH_3 \quad (2)$$

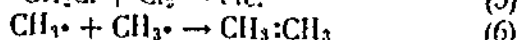
$$CH_3\cdot + Cl{:}Cl \rightarrow CH_3{:}Cl + Cl\cdot \quad (3)$$



$$Cl\cdot + H{:}CH_2Cl \rightarrow HCl + \cdot CH_2Cl \quad (4)$$



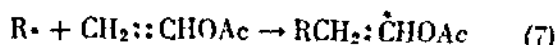
$$\cdot CH_2Cl + Cl_2 \rightarrow \text{etc.} \quad (5)$$



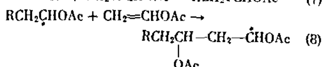
$$CH_3\cdot + CH_3\cdot \rightarrow CH_3{:}CH_3 \quad (6)$$

No reaction between chlorine and methane occurs in the dark, but irradiation of the mixture with light of the proper wavelength, short-wavelength visible light which is known to cause the dissociation of molecular chlorine into atoms (Eq. 1), causes a vigorous reaction that leads to the chlorinated methanes as products. The quantum yield, or the number of molecules of chlorine used up per photon absorbed by the system, is of the order of several thousand, and depends upon the conditions of the reaction. This means that dissociation of a molecule of chlorine into atoms initiates a chain reaction (steps 1–5) in which each step produces a new free radical, that is, a substance that has an odd number of extranuclear electrons, for example, $\cdot Cl$ or $\cdot CH_3$. These radicals then propagate the chain by attacking a molecule of substrate (CH_4 or Cl_2). Thus, a small number of initially produced radicals can lead to chemical change of a large number of reactant molecules. The chains are terminated by a number of processes, including those in which two radicals combine (Eq 6) Because the reaction chains are long, the products of such combination steps are formed in very low yields See PHOTOCHEMISTRY.

Vinyl polymerization, a type of reaction that is of immense industrial importance in the production of plastics, is another example of a homolytic chain process The reaction is initiated by free radicals ($R\cdot$) frequently produced by the decomposition of organic peroxides. The steps of the mechanism for the case of the vinyl acetate → polyvinyl acetate reaction are



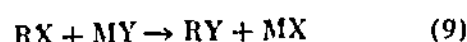
$$R\cdot + CH_2{::}CHOAc \rightarrow RCH_2{:}\dot{C}HOAc \quad (7)$$



$$RCH_2\dot{C}HOAc + CH_2{=}CHOAc \rightarrow RCH_2\underset{\displaystyle OAc}{\underset{|}{C}H}{-}CH_2{-}\dot{C}HOAc \quad (8)$$

Termination of the chains probably occurs mainly by combination (analogous to Eq 6) of polymeric radicals. See POLYMERIZATION

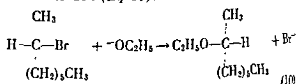
Heterolytic reactions. Equation (9) is a generalized scheme for a metathetical replacement or substitution, one of the most widely used reactions of synthetic organic chemistry In this general category are included such diverse processes as (1) conversions of halides or arenesulfonates to alcohols, ethers, thiols, or esters; (2) homologation of halides or arenesulfonates to nitriles or to substituted malonic or acetoacetic esters; (3) the Wurtz synthesis of alkanes; (4) the Gabriel synthesis of primary amines; (5) the alkylation of amines, (6) the preparation of alkanesulfonic acids and Bunte salts.



$$RX + MY \rightarrow RY + MX \quad (9)$$

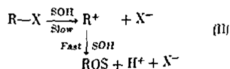
The heterolytic nature of these processes is strongly suggested by the facts that (1) MY is itself frequently an ionic reagent (an alkali hydroxide or cyanide, for example); (2) the reaction rates and products are sensitive to changes in reaction conditions (for example, added ionic species) that would not be expected to have marked effects upon homolytic processes; (3) the reaction rates and products are unresponsive to changes in conditions, such as light, sources of free radicals, and reagents that are known to react with free radicals that profoundly affect the course of homolytic processes.

Reactions of the type shown in Eq. (9) are often called nucleophilic displacements because the formation of RY involves formally the formation of a bond between R and an electron-rich, nucleus-seeking reagent (nucleophile) Y. In some cases, the reaction rate is proportional to the concentration in the solution of MY and RX. Such processes are interpreted as bimolecular, nucleophilic substitutions S_N2, in which both RX and Y are present in the transition state. Typically, these substitutions result in complete inversion of configuration (Walden inversion) if the reaction occurs at an asymmetric carbon (Eq 10).



$$H{-}\underset{(CH_2)_5CH_3}{\overset{CH_3}{C}}{-}Br + {}^{-}OC_2H_5 \rightarrow C_2H_5O{-}\underset{(CH_2)_5CH_3}{\overset{CH_3}{C}}{-}H + Br^- \quad (10)$$

Frequently, however, especially when R is a tertiary or α-phenylalkyl group, the reaction rate is essentially unaffected by the concentration of MY, and the reaction is termed a unimolecular, or S_N1, solvolysis. The reaction is not truly unimolecular because the solvent plays an important role and is undoubtedly involved in the transition state The mechanism shown in Eq (11) does not indicate the precise role of the solvent (SOH), because this varies from one case to the next, and is still imperfectly understood



$$R{-}X \xrightarrow[Slow]{SOH} R^+ + X^- ;\quad R^+ \xrightarrow[Fast]{SOH} ROS + H^+ + X^- \quad (11)$$

A carbonium ion, R^+, is postulated as a metastable intermediate in solvolytic reactions. The rates of appearance of H^+ (or X^-) parallel the stabilities of R^+ If X is attached to asymmetric center of R in optically active RX, ROS is usually extensively racemized Other products (RY) can also be formed from R^+ if Y^- is present See ADDITION REACTION; CONDENSATION REACTION; ELECTROPHILIC AND NUCLEOPHILIC REAGENT, HYDROXYLATION REACTION; STERIC EFFECT (CHEMICAL REACTIONS); SUBSTITUTION REACTION. [J A BE]

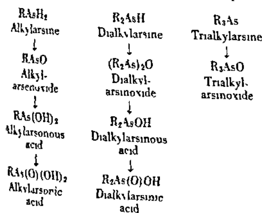
Bibliography: A. A. Frost and R. G Pearson, *Kinetics and Mechanism*, 1953; L. P. Hammett, *Physical Organic Chemistry*, 1940; J. Hine, *Physical Organic Chemistry*, 1956; C. K. Ingold, *Structure and Mechanism in Organic Chemistry*, 1953; C. Walling, *Free Radicals i*   *tion*, 1957

rganic reef

. reef consisting of the hard parts of organisms r of an organically constructed frame enclosing etrital particles, the hard parts of free-living or-anisms, and precipitated calcium carbonate Most rganic reefs are made of corals and associated rganisms, but some consist of lime-secreting algae, ydrozoans, annelids, oysters, or sponges Strictly peaking, a rocklike organic mass must be a men-ce to navigation before it can be classed as a reef. ctually, the term applies to any sizable organic minence that grows or once grew upward from he floor of a water body, ordinarily the sea *See* ATOLL; BARRIER REEF, CORAL REEF, FRINGING EEF; REEF. [P.E.C.]

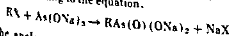
)rganoarsenic compound

A derivative of arsenic containing at least one or-ganic radical attached through carbon to the ar-enic atom by means of a covalent bond All or-anoarsenic derivatives can be derived schemati-ally by formal substitution or oxidation from the primary, secondary, and tertiary arsines Depend



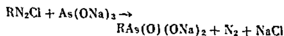
ing on the state of oxidation, organoarsenic com-pounds are derivatives of tri or pentavalent ar-senic. Although the known types of organoarseni-cals are *more* numerous than those of true organo-metallic compounds, they do not approach the num-ber of the known types of the organo derivatives of phosphorus which precedes arsenic in group V of the periodic system A special type of organoar-senic compounds which do not have an equivalent in the organophosphorus series are the arseno com-pounds, RAs═AsR

Preparation. The conversion of inorganic arsenic into organoarsenic compounds can be achieved readily in the aliphatic series by the Meyer reac-tion from trisodium arsenite and an alkylating agent according to the equation.



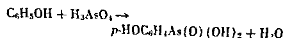
$$RX + As(ONa)_3 \rightarrow RAs(O)(ONa)_2 + NaX$$

The analogous alkylation of disodium alkylarson-ite, $RAs(ONa)_2$, yields sodium dialkylarsinate, $R_2As(O)ONa$, and sodium dialkylarsinite, R_2AsONa, leads similarly to trialkylarsine oxides. The ease with which derivatives of pentavalent arsenic are reduced to trivalent derivatives permits the synthesis of a large variety of organoarsenic com-pounds from sodium arsenite, and thus ultimately from arsenious oxide, As_2O_3. The alkylation of ar-senic trichloride, $AsCl_3$, with metalloorganic com-pounds, such as Grignard reagents and zinc or mercury dialkyls, presents an alternate method for the formation of a carbon-arsenic bond. Aromatic arsonic acids are conveniently accessible from the diazotized aromatic amines and sodium arsenite by the Bart reaction:



$$RN_2Cl + As(ONa)_3 \rightarrow RAs(O)(ONa)_2 + N_2 + NaCl$$

or by direct substitution of phenols or amines by arsenic acid according to the equation:



$$C_6H_5OH + H_3AsO_4 \rightarrow p\text{-}HOC_6H_4As(O)(OH)_2 + H_2O$$

Use. All organoarsensic compounds exhibit phys-iological activity. The high mammalian toxicity of the trivalent derivatives makes a number of *them potential* chemical warfare agents. Lewisite ($ClCH{=}CHAsCl_2$), Adamsite (phenarsazine chlo-ride), [illegible] $AsCl_2$), and [illegible] some of the [illegible] vesicants, ir-ritants, and sternutators Several of the less toxic derivatives have found use as chemotherapeutics among which Salvarsan (*arsphenamine* hydrochlo-ride), Atoxyl (sodium *arsanilate*) and Trypars-amide were of prime importance as the only means for combating certain infectious diseases be-fore the discovery of the antibiotics. *See* CHEMICAL WARFARE, INSECTICIDE, ORGANOPHOSPHORUS COM-POUND [F.W.H.]

Bibliography G T Morgan, *Organic Compounds of Arsenic and Antimony*, 1918, G. W. Raiziss and J L Gavron, *Organic Arsenical Compounds*, 1923.

Organometallic compound

One of a group of substances containing carbon-metal bonds This definition can be broadened to include all compounds of elements that possess metal-like properties and that are combined chemi-cally with one or more carbon atoms In this re-spect, metal alkoxides, chelates, salts of organic acids, and related compounds are *not* discussed in this article *See* CHELATION

Much of the early research in organometallic chemistry was concerned with the preparation and reactions of organozinc compounds, although the study of organic derivatives of other metals fol-lowed in rapid succession The pharmacological value of many organoarsenic and organomercury compounds stimulated much research on the syn-thesis and properties of these compounds The re-actions of Grignard reagents (*see* GRIGNARD REAC-TION) and of organolithium and organosodium com-pounds have found extensive use in synthetic or-ganic chemistry

Organometallic compounds have found many commercial applications. The use of tetraethyllead as a gasoline antiknock additive is a good example. *See* TETRAETHYLLEAD Organometallic compounds have played an important role in the development of silicone polymers, polyvinylchloride, and polyethylene.

A systematic classification of all known organometallic compounds is very difficult. For the sake of convenience, however, organometallic compounds can be divided into three main classes. The more electropositive metals of groups I and II of the periodic table form organometallic compounds that are nonvolatile, usually poorly soluble in organic solvents, and essentially ionic in nature. Lithium, beryllium, and magnesium compounds are much less ionic than those of the remaining metals, however. The metals (excepting transition metals) and metalloids of groups III, IV, V, and VI form organometallic compounds which are mainly volatile, soluble in organic solvents, and principally covalently bonded. The transition metals constitute a third main group, possessing bonding to aromatic or aromatic-type groups which is mainly of a special *d*-orbital, or "sandwich," type. *See* CHEMICAL BINDING.

Preparation from free metals. The reaction of metals or metalloids with organic halides is widely used in the synthesis of organometallic compounds



$$2M + RX \rightarrow RM + MX$$

Sodium, lithium, magnesium, and related metals react with organic halides in this manner to produce organometallic derivatives Silicon and germanium react readily in the gaseous phase with both alkyl and aryl halides.

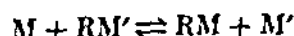
Another means of preparing organometallic compounds involves the reaction of highly reactive metals with hydrocarbons that contain active hydrogen atoms



$$RH + M \rightarrow RM + H$$

The preparations of organometallic derivatives of acetylene, triphenylmethane, and cyclopentadiene are examples of this reaction.

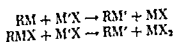
The displacement of a metal in an organometallic compound by a more reactive metal to produce a new organometallic compound has been widely used, although reactions of this type are usually reversible



$$M + RM' \rightleftharpoons RM + M'$$

Organometallic derivatives of the alkali metals, beryllium, magnesium, and aluminum have been prepared by this procedure

Preparation from metal salts. The reaction of or- [illegible]

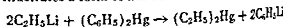


$$RM + M'X \rightarrow RM' + MX$$
$$RMX + M'X \rightarrow RM' + MX_2$$

In general, organic derivatives of more reactive metals react with metal salts of less reactive metals or metalloids by this procedure. Examples include the reactions of alkylsodium compounds, alkyllithium compounds, and Grignard reagents with halides of elements of groups III, IV, V, and VI. The recently discovered cyclopentadienyl derivatives of the transition metals can best be prepared in this manner.

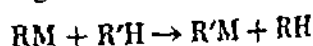
Metal salts react with aryldiazonium compounds in the presence of metals to give organometallic halides. Arylmercury and aryltin compounds have frequently been prepared in this way.

Other preparations. These reactions include methods by which an organometallic compound of a given metal is used to prepare other organometallic derivatives of the same metal The exchange reaction between ethyllithium and diphenylmercury illustrates a form of this reaction



$$2C_2H_5Li + (C_6H_5)_2Hg \rightarrow (C_2H_5)_2Hg + 2C_6H_5Li$$

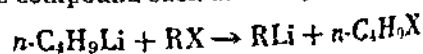
The reaction between an organometallic compound and an aromatic hydrocarbon possessing a reactive hydrogen atom is known as "metalation"



$$RM + R'H \rightarrow R'M + RH$$

The most commonly used "metalating agent," or RM compound, is *n*-butyllithium in ether solution Many other reactive organometallic compounds have also been used with success

A large number of organometallic derivatives of aromatic hydrocarbons have been prepared by the reaction of the corresponding aryl halides with a reactive compound such as *n*-butyllithium:



$$n\text{-}C_4H_9Li + RX \rightarrow RLi + n\text{-}C_4H_9X$$

Compounds of group I elements. The alkali metals form organometallic compounds that possess high reactivity. Most alkyl derivatives are colorless solids, except the higher alkyls of lithium, which are liquids. Many aryl derivatives are solids, and have intense colors that depend on the number of aromatic and conjugated groups in the molecule With the exception of the alkyllithium derivatives, organoalkali compounds are characterized by insolubility in most organic solvents and little tendency to vaporize or melt without decomposition Such properties indicate a large degree of ionic character in the bonding between the metal and organic group Because of the small size and high polarizing power of the lithium ion, alkyl derivatives of lithium are largely covalently bonded

Because of the high reactivity of organoalkali compounds, they must be prepared and used in an inert atmosphere that is free of oxygen, moisture and carbon dioxide. Their high reactivity makes organoalkali compounds very useful in synthetic organic chemistry.

Compounds of group II elements. All elements of group II except radium form organometallic compounds, although no simple compounds of calcium, strontium, or barium have yet been isolated A remarkable gradation in physical and chemical properties is observed for organometallic compounds of this group. Nearly [illegible] these compounds except those of [illegible] le in the

presence of oxygen and water. Alkylberyllium compounds are volatile, although highly associated liquids; an exception is dimethylberyllium. Arylberyllium compounds are solids and are somewhat more stable.

Two main types of organomagnesium compounds are known, alkyl- and arylmagnesium compounds, R_2Mg, and the organomagnesium halides or Grignard reagents, RMgX. The latter compounds are the most widely known organometallic derivatives of group II elements and have found extensive use in organic syntheses. Alkyl- and arylmagnesium compounds are white crystalline solids. They are practically nonvolatile and are insoluble in hydrocarbon solvents. Although these compounds are saltlike, the bonding between the magnesium and the organic group is believed to be largely covalent.

Organozinc and organocadmium compounds were among the first organometallic compounds to be isolated. For many years organozinc compounds were used for synthetic purposes until they were superseded by the more convenient Grignard reagents. Alkylzinc and alkylcadmium compounds are volatile, unassociated liquids; aryl derivatives are white solids with sharp melting points. *See* REFORMATSKY REACTION.

Many organomercury compounds of the type R_2Hg and RHgX have been prepared. The reactivities of these compounds are so low that they are unaffected by water and air. Dialkylmercury compounds are mostly distillable liquids; diarylmercury derivatives are crystalline solids. The bonding in these compounds is essentially covalent. Alkyl- and arylmercury compounds of the type RHgX are stable crystalline solids, their properties depend largely on the nature of the group X.

Compounds of group III elements. Most alkyl and many aryl derivatives of the group III elements are spontaneously flammable in the presence of air. The alkylboron compounds are colorless liquids, the arylboron compounds are crystalline solids. Alkyl- and arylboric acids, RB(OH)... are impor... compounc... with aton... electron-donor groups...

Alkylalum... of extreme... shown to po... the corresp... minum com... and important industrial applications as polymerization catalysts (see POLYOLEFIN RESINS). Arylaluminum compounds are crystalline solids and are also highly reactive.

A few alkyl and aryl derivatives of gallium, indium, and thallium are known. In general, these compounds exhibit a decreasing order of reactivity, especially when compared to the corresponding organoaluminum compounds. A recent notable development has been the isolation of a monovalent organoindium compound, cyclopentadienylindium, C_5H_5In.

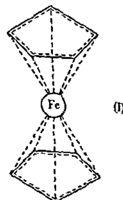
Compounds of group IV elements. The group IV elements silicon, germanium, tin, and lead form organometallic compounds of the type R_4M. All compounds of this type are covalently bonded, although the carbon-metal bond becomes weaker and more polar as the metal becomes more metallic. A special feature of elements of this group is their ability to form metal-metal bonds...

wide temperature ranges. Aryl derivatives are high-melting solids. Both alkyl and aryl derivatives are usually unattacked by water and air at room temperature, and are soluble in hydrocarbon solvents.

Compounds of group V and group VI elements. Organic derivatives of phosphorus, arsenic, antimony, and bismuth are well known, although it is debatable whether compounds of the first two elements can be classified as truly organometallic. Group V elements form organometallic compounds of both R_3M and R_5M types and also many mixed organometallic hydrides, halides, and hydroxides. Alkyl derivatives of the type R_3M are liquids of decreasing stability in the case of the more metallic members. Recently a number of pentaphenyl derivatives of the type $(C_6H_5)_5M$ have been reported in which all phenyl groups are covalently bonded, although the fifth phenyl group is not strongly held.

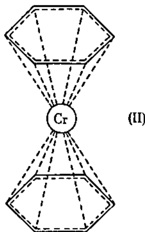
Both selenium and tellurium in group VI form alkyl and aryl derivatives of the type R_2M. In some aspects these can be considered to be organometallic compounds. The alkyl derivatives are colorless liquids and the aryl derivatives are low-melting solids, all possess extremely unpleasant odors. A number of diselenides and ditellurides of the type RM—MR are also known.

Compounds of the transition metals. Until recently it was thought that transition metals were incapable of forming organometallic compounds, although many scattered claims have appeared throughout the literature. In 1951 a new class of organometallic derivatives of transition metals was discovered. The compounds in this class are different in most respects from the organometallic compounds described in the preceding sections. The first compound of this type to be isolated was dicyclopentadienyliron, commonly called ferrocene (I).



(I)

This ...



atom to form a symmetrical, extremely stable structure. It has been suggested that the iron atom is bonded to each cyclopentadienyl ring by a delocalized covalent bond Dicyclopentadienyliron has been shown to possess highly aromatic properties and has thus opened up an entirely new field of organic chemistry.

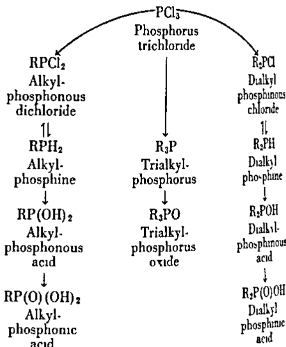
Since the initial discovery of dicyclopentadienyliron, the isolation of cyclopentadienyl derivatives of most of the other transition metals has followed in rapid succession Many of these compounds possess the sandwich structure of the iron analog, although essentially ionic and covalent types are also known

A related class of compounds which possess a similar symmetrical structure has also been recently reported. These compounds can be illustrated by the most stable member, dibenzenechromium (II). An organometallic compound possessing a phenyl-titanium bond, phenyltitanium triisopropoxide, $C_6H_5Ti(OC_3H_7)_3$, is also known [M D R.]

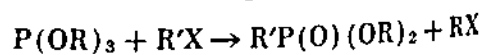
Bibliography: E G. Rochow, D. T. Hurd, and R. N. Lewis, *The Chemistry of Organometallic Compounds*, 1957.

Organophosphorus compound

One of a series of derivatives of phosphorus which have at least one organic (alkyl) group attached to the phosphorus atom by direct linkage to a carbon atom The alkylphosphines (RPH_2), dialkylphosphines (R_2PH), and trialkylphosphorus (trialkylphosphines, R_3P) can be regarded formally as the parent compounds of all organophosphorus compounds. Formal substitution of the hydrogen of the phosphines by monovalent atoms or groups leads to a number of basic structures derived from trivalent phosphorus, and formal addition of bivalent oxygen leads from the alkyl- and dialkylphosphines to organophosphorus acids and from the trialkylphosphines to their oxides (see chart) Similar additions of other bivalent atoms or groups, such as sulfur, selenium, or :NH, lead to related structures containing phosphorus in its tri- or pentavalent state. Considering the large number of organic groups which may be linked by a direct bond to the phosphorus, it is obvious that the number of theoretically possible organophosphorus compounds is practically unlimited Only a comparatively small number of these possibilities has been realized.



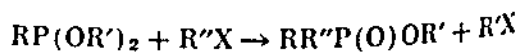
Methods of preparation. A number of organophosphorus compounds containing an unsubstituted or substituted aliphatic group directly attached to the phosphorus atom has been prepared by the Michaelis-Arbuzov reaction from trialkyl esters of phosphorous acid (trialkyl phosphites) and aliphatic halides according to the general equation



$$P(OR)_3 + R'X \rightarrow R'P(O)(OR)_2 + RX$$

This reaction involves a change of valency of the phosphorus from the trivalent to the pentavalent state. The resulting dialkyl alkylphosphonates are versatile starting materials for the conversion by standard chemical methods to a variety of other classes of compounds. A frequently used alternate preparative method for dialkyl alkylphosphonates consists of the alkylation of the sodium derivatives of dialkyl hydrogen phosphonates, $(RO)_2P(O)H$, with aliphatic halides. The direct alkylation of phosphorus halides with organometallic compounds finds only limited application for the synthesis of compounds containing one carbon-to-phosphorus linkage, but is useful for the preparation of trialkylphosphines

Derivatives with two aliphatic groups attached to pentavalent phosphorus are also accessible by the Michaelis-Arbuzov reaction from dialkyl alkylphosphonites:



$$RP(OR')_2 + R''X \rightarrow RR''P(O)OR' + R'X$$

or by alkylation of the sodium derivative of the appropriate monoesters $RP(O)(OR')H$.

Since the preparative methods involving alkyl halides depend on the reactivity of the aliphatic halide, aromatic organophosphorus compounds cannot be obtained in an analogous manner from aryl halides The preferred method for the preparation of the aromatic derivatives is the introduction of a $—PCl_2$ group into aromatic hydrocarbons by means of phosphorus trichloride ... -anhydrous aluminum chloride as a catal... ...ing arylphos-

phonous dichlorides can react further with the aromatic hydrocarbon to yield the diarylphosphinous chlorides, Ar_2PCl, which are always obtained in addition to the phosphonous dichlorides, but can be made the main product of the reaction by proper choice of the reaction conditions

Organophosphorus derivatives containing a P—H bond can be added across activated olefinic and carbonyl double bonds to yield with the formation of C—P bond adducts with a more complex alkyl group attached to the phosphorus Dialkyl esters of phosphorous acid, $(RO)_2P(O)H$, lead by this reaction to dialkyl alkylphosphonates The addition of P—H derivative to the carbonyl group of an aldehyde results in the formation of α-hydroxy compounds susceptible to further modification of the hydroxyl group by standard chemical reactions

Uses. In contrast to the wide commercial applications of organic derivatives of phosphorus acids, little use has been found for organophosphorus compounds, although some have been used as polymerization catalysts, lubricant additives, flameproofing agents, plant growth regulators, solvents for the solvent extraction of uranium, and insecticides The high mammalian toxicity exhibited by some methylphosphonic acid derivatives, which are extremely potent inhibitors of the enzyme cholinesterase limits the usefulness of a large number of related, although much less toxic compounds for commercial exploitation because of the potential health hazards During World War II the Germans developed several highly toxic organophosphorus compounds for use as chemical-warfare agents The most potent of these nerve gases were isopropyl methylphosphonofluoridate, $CH_3P(O)[OCH(CH_3)_2]F$ (Sarin or Trilon 46), the homologous pinacolyl ester $CH_3P(O)[OCH(CH_3)C(CH_3)_3]F$ (Soman), and ethyl phosphorodimethylamidocyanidate $(CH_3)_2NP(O)(OC_2H_5)CN$ (Tabun). See CHEMICAL WARFARE; INSECTICIDE; PHOSPHORUS

[F.W.H.]

Bibliography: G. M. Kosolapoff, *Organophosphorus Compounds*, 1950

Organosulfur compound

One of a group of substances which contain both carbon and sulfur The elements oxygen nitrogen, the halogens, and phosphorus are also often present Thousands of such compounds are well known

Many medicinal and natural products are organosulfur compounds, for example, the penicillins sulfa drugs, insulin, and amino acids such as cysteine and methionine The utilizations of organosulfur compounds involve enormous commercial developments in medicinals detergents, sulfide rubbers and polymers, sulfur dyes, solvents, and agricultural chemicals (herbicides fungicides, and insecticides) Besides their practical uses, studies of organosulfur compounds helped to establish fundamental theory on the nature of valence bond ing, of molecular geometry, and of reaction mechanisms of organic chemistry

The most common classes of organosulfur compounds are: (1) mercaptans, RSH, also called thiols or sulfhydryl compounds, such as methyl mercaptan, CH_3SH or *tert*-butylmercaptan, $(CH_3)_3CSH$; (2) thiophenols, ArSH, for example, thiophenol, C_6H_5SH; (3) disul[illegible] [illegible]hatic, aro[illegible] [illegible]des, includ[illegible] [illegible]ethyl ethyl [illegible]

$$H_2C \!-\!\!-\! CH_2 \quad \text{(ring closed through S)}$$

(5) sulfenyl chlorides, RSCl, for example, trichloromethanesulfenyl chloride, Cl_3CSCl; (6) sulfoxides,

$$R-\overset{O}{S}-R$$

for example, dimethyl sulfoxide, an interesting solvent; (7) sulfones,

$$R-\overset{O}{\underset{O}{S}}-R$$

for example, bis-(*p*-aminophenyl) *sulfone*, $NH_2C_6H_4SO_2C_6H_4NH_2$, of interest in the treatment of tuberculosis; (8) sulfonyl *chlorides*, RSO_2Cl, for example, *p*-toluenesulfonyl *chloride*, $p\text{-}CH_3C_6H_4SO_2Cl$, (9) sulfonamides, $R-SO_2NH_2$, *for* example, benzenesulfonamide, $C_6H_5SO_2NH_2$; (10) thioaldehydes and *thioketones*, $RC(H)=S$ and $R_2C=S$, which generally exist as trimers, *for* example, trithioacetaldehyde, (11) *sulfinic* acids, RSO_2H, (12) sulfonic acids, RSO_3H; (13) *esters* of sulfonic acids, RSO_2OR'; (14) thiol esters,

$$R\overset{O}{\overset{\|}{C}}-SR'$$

and thio acids,

$$R\overset{O}{\overset{\|}{C}}-SH$$

A large group of organosulfur compounds belongs to the heterocyclic series In these compounds the sulfur atom is part of a ring system, and many others belong to groups not mentioned above, such as thiocarbonates, dithio acids, sulfimides, sulfamic acids, thioureas, sulfonium salts, mercaptals, mercaptols, organic thiocyanates, and isothiocyanates.

Selenium follows sulfur in the sixth main family of the periodic system and hence has many similarities to sulfur Organoselenium compounds are also relatively well known; many of them are analogous to the sulfur compounds. See HETEROCYCLIC COMPOUNDS; PETROLEUM PROCESSING; SULFONAMIDE

[N.K.]

Bibliography T. W. Campbell, H. G. Walker and G. M. Coppinger, Some aspects of the organic chemistry of selenium, *Chem. Revs*, 50:279, 1952; L. Fieser and M. Fieser, *Organic Chemistry*, 3d ed., 1956; H. Gilman, *Organic Chemistry*, vol. 1, 2d ed., 1943; R. E. Kirk and D. F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vols. 8, 13, and 14, 1954; E. E. Reid, *Organic Chemistry of Bivalent*

Sulfur, vol. 1, 1958; C. M. Suter, *The Organic Chemistry of Sulfur*, *Tetracovalent Sulfur Compounds*, 1944.

Orifice

An opening in a wall through which fluid flows, the thickness of the wall being less than a fifth of the area of the opening and the approach curvature being negligible. An orifice serves basically to control or to meter the rate of fluid flow. It may be used as an instrument for the measurement of fluid flow or as an element in a machine to limit the flow of fluid, including oils, air, gases, steam, and other vapors. Common shapes are sharp-edge, with bevel facing out, and square (Fig. 1). Often, fuel enters a chamber through an orifice, as in a carburetor.

In both liquid and gas devices, an orifice acts to convert potential energy to kinetic energy. For liquids, the potential energy is usually the pressure head; for gases, the potential energy includes the pressure differential and the temperature available for conversion into velocity. Theoretical velocity v of discharge of liquid initially at rest from an orifice is $v = (2gh)^{1/2}$ where g is gravitational constant and h is pressure head (see TORRICELLI'S THEOREM). For an ideal incompressible liquid of specific weight γ lb/ft^3 flowing through an orifice of area a in^2 whose discharge coefficient, including velocity of approach, is K, the rate of flow w_s in lb/sec is

$$w_s = Ca\gamma\sqrt{2gh}$$

where g is the acceleration of gravity in ft/sec^2 and h is the effective differential head or change in pressure in feet measured across the orifice.

For liquid measurement, an orifice is arranged as in Fig. 2. The liquid passes through the orifice, experiencing a pressure drop which is transmitted to a liquid-filled manometer, differential pressure gage, or to a recording instrument. The flow rate is calculated from the change in pressure, the area of the orifice, and the foregoing relation for velocity. *Figure 2 also shows, as the rise of liquid in manometers* along the pipe, the effect of the orifice on pressure. The region where the pressure is lowest, with the stream lines closest together, is called the vena contracta.

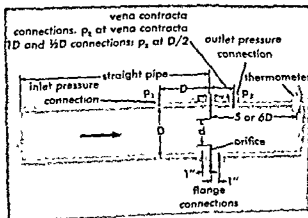


Fig. 1. Dimensions of orifice installation in pipe of diameter D.

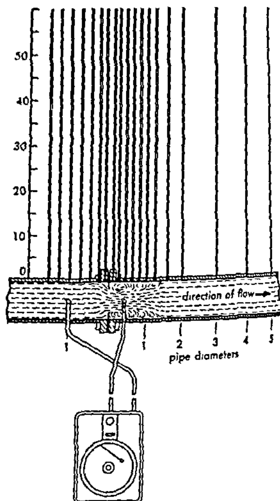


Fig 2 Diagrammatic representation of fluid flow through a thin-plate orifice.

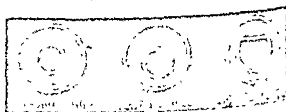


Fig. 3. Commonly used orifice shapes.

Orifices for flow instrumentation are clamped between pipe flanges. They may have a concentric hole, although for self-cleaning the hole may be off center with one edge flush with the bottom of the pipe or with a segmented hole equal to the inside diameter of the pipe (Fig. 3)

As a differential pressure producer, the orifice is the most widely used device. More test data are available on the thin plate orifice than on other type flow meters, thus it is chosen despite its virtual lack of pressure recovery (see VENTURI TUBE). The orifice also has a low discharge coefficient, which is a disadvantage in instrumentation but an advantage in flow control For greatest accuracy, corrections are made for approach velocity and for temperature differential. See FLOW MEASUREMENT. [R E SP.]

Bibliography: Am Soc. Mech. Engrs , *Flow Measurement*, chap. 4, pt. 5, suppl. to ASME Power Test Codes, PTC 19 5:4; 4–195 · ..M. Considine (ed), *Process Instruments a* ... *Handbook* 1957

Oriole

Any of 30 species of American songbirds comprising the genus *Icterus*, family Icteridae. Six of these species are found in the United States. Male orioles are brilliantly colored birds, usually showing a combination of black with yellow, orange, or chestnut. Females are usually olive and yellow, in subdued tones. They build elaborate woven nests, suspended from twigs. Best known is the Baltimore



The Baltimore oriole, *Icterus galbula*, length to 8⅛ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

oriole, *I. galbula*, a brilliant black and orange bird, 7–8 in. long, found over most of the eastern United States. Bullock's oriole, *I. bullocki*, is a similar western species. The orchard oriole, *I. spurius*, is a smaller bird marked in chestnut and black, also common over much of the eastern United States. See PASSERIFORMES. [J.D.B.]

Orion

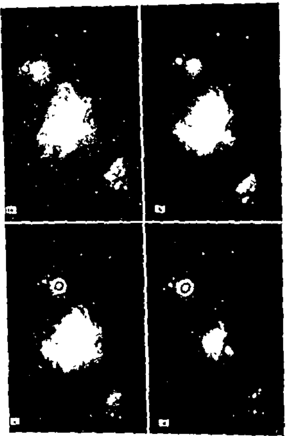
The Warrior, in astronomy, and undoubtedly the finest of all constellations in the sky. Orion is a winter group near the celestial equator. Four of the most prominent stars form a huge crude rectangle. The group is pictured as the figure of a warrior, holding a shield with his left hand and swinging a club with his raised right arm ready to strike the [illegible]

[illegible]

[illegible] three bright stars in a straight line in the middle of the rectangle represent the warrior's belt. The center star is Alnilam. These four stars just cited are all navigational stars. Three faint stars below the belt form the Sword of Orion, the middle one of which is actually four very hot stars, the Trapezium, embedded in the Great Nebula of Orion. See CONSTELLATION. [C.S.Y.]

Orion nebula

A cloud of gas and solid grains which shines because of radiation received from high temperature (~30,000°K) stars embedded within it, as illustrated. The inner region has a radius of about 4′



Orion nebula, NGC 1976, photographed through Curtis Schmidt telescope at the University of Michigan Observatory. Image produced by (a) green nebular lines of oxygen [O III], (b) ultraviolet region, chiefly oxygen [O II], (c) red hydrogen line, and (d) continuum emission

(which corresponds to about 0.6 parsecs) and a mass on the order of 10 solar masses. The electron density is about 30,000/cm³ near the brightest part (called the Trapezium) and about 300/cm³ in the outer portion. The electron temperature is near 10,000°K. The distance of the Orion nebula is in the neighborhood of 500 parsecs. Radio observations show that the Orion nebula and entire complex of bright stars, gas, and grains are surrounded [illegible]

[illegible] NEBULA, GASEOUS. [L.H.A.]

Ornamental plant

Any plant cultivated essentially for decorative purposes. Ornamental herbaceous or nonwoody plants include a wide variety of species such as carnation, chrysanthemum, gladiolus, tulip, iris, and sweet pea. Their commercial value in the United States probably exceeds $1,000,000,000 annually. Herbaceous ornamentals are grown in open fields or under the protection of lath, plastic sheeting, cloth, or glass structures. Nonglasshouse production is concentrated in the southern United States and on the Pacific Coast, especially in Florida and Cali-

fornia. Glasshouse production occurs mainly in the northern and eastern urban areas.

Ornamental shrubs include numerous wild and cultivated perennial woody plants, such as rose, azalea, camellia, lilac, privet, boxwood, juniper, cedar, and arborvitae. Ornamental shrubs are generally produced in the open. Protection is usually limited to lath covers in areas of intense sunlight and high temperatures and to windbreaks where desiccating air currents create a problem. The annual value of ornamental shrubs amounts to many millions of dollars. *See* FLORICULTURE; FOREST TREE NURSERY MANAGEMENT, LANDSCAPE ARCHITECTURE; TREE. [J.F.F.]

Diseases of ornamental herbs. A serious and universal disease, commonly called damping-off (killing of plants before they emerge from the soil or shortly thereafter), is caused by *Rhizoctonia solani*, *Thielaviopsis basicola*, and species of *Pythium*, *Phytophthora*, and *Fusarium* The same fungi may cause root rot of older plants. Strains of *Fusarium oxysporum* parasitize xylem vessels (water-conduction tissue) of many plants causing desiccation and resulting in severe losses Examples are aster wilt, carnation wilt, and gladiolus yellows (wilt). This fungus is unique because it contains strains, or forms, which have specific and limited host ranges, for example, the strain which attacks china aster cannot infect gladiolus. In contrast, *Verticillium albo-atrum*, a vascular wilt parasite especially damaging to chrysanthemum, has a wide, nonspecific host range.

Because cut flowers, such as carnation, snapdragon, and gladiolus, must be free of blemishes, diseases of flower parts often result in heavy losses. Gray mold, caused by *Botrytis cinerea*, is the most common disease of this type Control is difficult because the flowers have little inherent resistance, the fungus thrives in cool wet conditions, making application of fungicides difficult, and sometimes fungicides injure flowers and cannot be used

Viruses may kill plants or mottle or spot leaves and flowers (calla spotted wilt on leaves, sweet pea mosaic, tulip break), distort flowers (delphinium petals infected with aster yellows revert to sepal and leaflike shapes); curl, cup, or distort leaves (geranium leaf curl, curly top of petunia), or stunt plants (chrysanthemum stunt).

Of increasing importance are physiological diseases caused by nonparasitic agents Air-pollution injury resulting from smog and illuminating and industrial gases, and salinity, nutritional, and cultural damage are examples of physiological diseases which often limit successful plant growth.

Controls include obtaining and maintaining pathogen-free planting material; sterilizing soil, benches, pots, tools, and other equipment by steam or chemicals; observing strict sanitary precautions to keep the plants healthy; and applying fungicides and insecticides to control air-borne pathogens. *See* FUNGISTAT AND FUNGICIDE; HERBICIDE.

Diseases of ornamental shrubs. Many diseases of ornamental shrubs are restricted by the host environment; for example, black spot of rose (*Diplocarpon rosae*) is a serious disease except in the arid Southwest and California, and azalea flower blight (*Ovulinia azaleae*) is most serious in regions of high humidity and high temperatures.

Root-rotting organisms (*Rhizoctonia solani*, *Phytophthora cinnamomi*, and *Armillaria mellea*) often cause heavy losses. They also cause rots of cuttings, seedlings, and other propagative material. Crown gall (*Agrobacterium tumefaciens*), a bacterial infection which results in root galls, is frequently encountered.

Species of rusts and powdery mildews are generally specific for individual hosts and are usually not capable of parasitizing other plants. Rust diseases are important because they not only affect ornamental plantings directly (rose rust), but also indirectly, because several shrubs are alternate host plants for cereal rusts (buckthorn-oat rust, barberry-wheat stem rust). For this reason some ornamental plants are prohibited in many cereal producing areas. Other leaf-spotting diseases, such as black spot of rose, are quite troublesome and result in stunted and unsightly plants.

Of the stem diseases, the canker-producing fungi (rose and gardenia cankers, fire blight of cotoneaster and pyracanthus) are frequently important. These infections result in death of plant parts at and beyond the point of infection

Flower blights of azalea (*Ovulinia azaleae*) and camellia (*Sclerotinia camelliae*) are diseases which infect only flowers. The parasites exist between flowering periods as sclerotia (hard, resistant reproductive bodies). Control is difficult, because the sclerotia are lodged in soil debris, and they are difficult to reach with fungicides

Virus diseases of shrubs are not generally considered important and probably for this reason not so much research has been done with them However, on roses several viruses have been described, and one is responsible for color break of flowers on some camellia varieties.

Control of many diseases of perennial shrubby plants depends upon maintenance of optimum growing conditions, good drainage and aeration, systematic fertilization, and correct applications of fungicides and insecticides *See* INSECTICIDE; PLANT DISEASE; PLANT VIRUS. [D.E.M.]

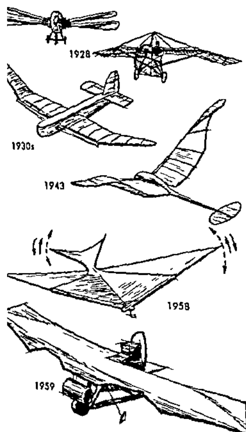
Ornithischia

An extinct order of reptiles belonging to the subclass Archosauria. The group includes those dinosaurs which had a birdlike pelvis and a predentary bone in the lower jaw. All ornithischian dinosaurs were plant-eaters. *See* ARCHOSAURIA; DINOSAUR [J.T.G.]

Ornithopter

An aircraft whose wings beat like those of a bird, bat, or insect (*see* FLIGHT). Models are recorded as early as 400 B.C. (Archytas), and for more than a century have been flown successfully, as illustrated.

Conditions for efficient flight. The forces required to lift and propel heavier-than-air aircraft



ɔresentative flying models of ornithopter show the ny possible variations

: produced by *reaction* An upward or forward ıction occurs when streams of air are directed wn or to the rear by moving parts of the aircraft ese moving parts may be propellers, fans, or ating wings The fans or propellers may be open in the conventional aircraft or helicopter, or en-ısed as in a jet engine or ducted fan

The fundamental laws of physics show that a ıen reactive force will be produced with the least ·s of energy when the mass flow of the fluid eam is large and its velocity small Reactive rce is proportional to the product of the fluid ıss moved per unit of time and its acquired veloc-·. Energy is proportional to the product of mass d velocity squared, divided by two Thus, for a ren energy loss, the force obtained is inversely oportional to the velocity of the stream

To the aircraft designer, this means that in low eed and hovering the least engine power will be quired when the cross-sectional area of the air-ream is large If an aircraft which fulfills this at condition can then also cruise so that its mov-g propelling surfaces attain only a small velocity rtement over the basic velocity of the aircraft ıd are functionally stall-resistant then the over l performance of this aircraft for a given engine ıwer will be maximum. Among existing and pro-ısed designs, only the ornithopter seems to meet ese requirements. There appears to be no aerody ımic limit to the speed which it might attain.

Flight tests. An Italian model flown in the 1930s weighed nearly 50 lb and was powered by a ½-horsepower (hp) air motor. About 1890 O Lilienthal demonstrated a static lift of more than 70 lb/hr. Other tests show neither excessive power losses nor insurmountable structural problems in beating wings

Numerous ornithopter designs have been tried Some have resembled birds with jointed, variable-area folding wings Wings have had simple surfaces like the bat, feathers, slots, and valves have also been tried Single-spar leading edge wings with trailing flexible membranes have been popular in models. Various forms of motive power and power transmission, including indirect vibratory excitement of an elastically tuned wing system, have been used A number of full-scale research and development projects are now active [J.L.C F.]

nautical Sciences, 1952.

Orogeny

The process or processes of mountain formation, especially the intense deformation of rocks by folding and faulting which, in many mountainous regions, has been accompanied by metamorphism, invasion of molten rock, and volcanic eruption Interpreted literally, the term orogeny includes both the structural deformation of rocks and the uplift that raised the mountains as a topographic form —two processes that have not always acted simultaneously As ordinarily used, however, orogeny signifies the deformation of the rocks, whether this happened during or before the mountain uplift. Orogeny is sometimes used in contrast to epeirogeny, the process of continent formation, a term signifying the more widely extended but gentler disturbances of the earth's crust by which broader features such as continents and ocean basins have risen or sunk This distinction between the processes that form mountains and those that move con-

b

o

tl

... mountain systems—mountains now worn down to their roots and buried beneath a cover of younger and relatively undisturbed sedimentary rocks The continents may thus have grown to their present size by orogenic processes operating now here, now there, throughout the long course of earth history *See* DIASTROPHISM.

GENERAL RELATIONSHIPS

Each mountain range has its own individual characteristics. Yet field studies by geologists have disclosed marked similarities in the geometrical pattern of rock masses and in the deformational history of many of the great mountain systems of the earth Not all the elements of pattern and history herein discussed are found in every mountain range, but each feature is exhibited in varying degree in all larger mountain systems.

fornia. Glasshouse production occurs mainly in the northern and eastern urban areas.

Ornamental shrubs include numerous wild and cultivated perennial woody plants, such as rose, azalea, camellia, lilac, privet, boxwood, juniper, cedar, and arborvitae. Ornamental shrubs are generally produced in the open. Protection is usually limited to lath covers in areas of intense sunlight and high temperatures and to windbreaks where desiccating air currents create a problem The annual value of ornamental shrubs amounts to many millions of dollars. *See* FLORICULTURE, FOREST TREE NURSERY MANAGEMENT; LANDSCAPE ARCHITECTURE; TREE [J.F.F.]

Diseases of ornamental herbs. A serious and universal disease, commonly called damping-off (killing of plants before they emerge from the soil or shortly thereafter), is caused by *Rhizoctonia solani*, *Thielaviopsis basicola*, and species of *Pythium*, *Phytophthora*, and *Fusarium*. The same fungi may cause root rot of older plants. Strains of *Fusarium oxysporum* parasitize xylem vessels (water-conduction tissue) of many plants causing desiccation and resulting in severe losses Examples are aster wilt, carnation wilt, and gladiolus yellows (wilt). This fungus is unique because it contains strains, or forms, which have specific and limited host ranges; for example, the strain which attacks china aster cannot infect gladiolus. In contrast, *Verticillium albo-atrum*, a vascular wilt parasite especially damaging to chrysanthemum, has a wide, nonspecific host range.

Because cut flowers, such as carnation, snapdragon, and gladiolus, must be free of blemishes, diseases of flower parts often result in heavy losses Gray mold, caused by *Botrytis cinerea*, is the most common disease of this type Control is difficult because the flowers have little inherent resistance; the fungus thrives in cool wet conditions, making application of fungicides difficult, and sometimes fungicides injure flowers and cannot be used.

Viruses may kill plants or mottle or spot leaves and flowers (calla spotted wilt on leaves, sweet pea mosaic, tulip break); distort flowers (delphinium petals infected with aster yellows revert to sepal and leaflike shapes); curl, cup, or distort leaves (geranium leaf curl, curly top of petunia); or stunt plants (chrysanthemum stunt).

Of increasing importance are physiological diseases caused by nonparasitic agents. Air-pollution injury resulting from smog and illuminating and industrial gases, and salinity, nutritional, and cultural damage are examples of physiological diseases which often limit successful plant growth

Controls include obtaining and maintaining pathogen-free planting material; sterilizing soil, benches, pots, tools, and other equipment by steam or chemicals; observing strict sanitary precautions to keep the plants healthy; and applying fungicides and insecticides to control air-borne pathogens *See* FUNGISTAT AND FUNGICIDE; HERBICIDE.

Diseases of ornamental shrubs. Many diseases of ornamental shrubs are restricted by the host environment; for example, black spot of rose (*Diplocarpon rosae*) is a serious disease except in the arid Southwest and California, and azalea flower blight (*Ovulinia azaleae*) is most serious in regions of high humidity and high temperatures

Root-rotting organisms (*Rhizoctonia solani*, *Phytophthora cinnamomi*, and *Armillaria mellea*) often cause heavy losses. They also cause rots of cuttings, seedlings, and other propagative material. Crown gall (*Agrobacterium tumefaciens*), a bacterial infection which results in root galls, is frequently encountered.

Species of rusts and powdery mildews are generally specific for individual hosts and are usually not capable of parasitizing other plants. Rust diseases are important because they not only affect ornamental plantings directly (rose rust), but also indirectly, because several shrubs are alternate host plants for cereal rusts (buckthorn-oat rust, barberry-wheat stem rust). For this reason some ornamental plants are prohibited in many cereal producing areas. Other leaf-spotting diseases, such as black spot of rose, are quite troublesome and result in stunted and unsightly plants

Of the stem diseases, the canker-producing fungi (rose and gardenia cankers, fire blight of cotoneaster and pyracanthus) are frequently important. These infections result in death of plant parts at and beyond the point of infection.

Flower blights of azalea (*Ovulinia azaleae*) and camellia (*Sclerotinia camelliae*) are diseases which infect only flowers. The parasites exist between flowering periods as sclerotia (hard, resistant reproductive bodies) Control is difficult, because the sclerotia are lodged in soil debris, and they are difficult to reach with fungicides.

Virus diseases of shrubs are not generally considered important and probably for this reason not so much research has been done with them However, on roses several viruses have been described, and one is responsible for color break of flowers on some camellia varieties

Control of many diseases of perennial shrubby plants depends upon maintenance of optimum growing conditions, good drainage and aeration, systematic fertilization, and correct applications of fungicides and insecticides. *See* INSECTICIDE; PLANT DISEASE, PLANT VIRUS. [D.E.M.]

Ornithischia

An extinct order of reptiles belonging to the subclass Archosauria The group includes those dinosaurs which had a birdlike pelvis and a predentary bone in the lower jaw All ornithischian dinosaurs were plant-eaters. *See* ARCHOSAURIA; DINOSAUR [J.T.G.]

Ornithopter

An aircraft whose wings beat like those of a bird, bat, or insect (*see* FLIGHT). Models are recorded as early as 400 B.C. (Archytas), and for more than a century have been flown successfully, as illustrated

Conditions for efficient flight. The forces required to lift ... el heav ... aircraft

...ent years in the study of a region indicate that ...y of the younger mountain chains have under... a net shortening of structure, apparent or of 50–100 miles or even more. It seems likely the far more numerous older mountain systems, ...h are now eroded down to their roots and thus ...accessible to similar detailed studies, have also ...ergone comparable amounts of this shortening ...aggregate amount of apparent structural short...g in all mountain systems, young and old, must ...ery large

...ock metamorphism and igneous intrusion. In ...central core areas of the more intensely de...med mountains, the minerals of which the origi... sedimentary rocks were composed have been ...tamorphosed by high temperatures and pressures ...l by hot solutions so that the rocks are now ...artzite, slate, schist, gneiss, and marble Exam...s are found on the southeastern slope of the ...ppalachians; in the old Caledonian mountains of ...otland, Norway, and Sweden, in the southern ...lps; and in the central Himalayas

In the central core areas of other mountain sys...ms, great masses of molten rock, or magma, ...ve risen from deep in the crust and invaded the ...eformed rocks, in some places greatly altering ...heir original texture and composition In many ...laces, these molten masses cooled underground ...nd, having been uncovered by later stream ero...sion are now represented by extensive exposures ...f granitic and other igneous rocks The great batholiths of the Sierra Nevada, Idaho, and British Columbia in North America and of the Andes in South America are examples In other places, the magma broke through to the surface and formed volcanoes or poured out in great lava fields—for example, the Cascade Range and the Columbia River lava fields of northwestern United States and the great Deccan traps (plateau steps) of India

Island arcs. The above generalizations about mountains and mountain building are based upon relations observed on land. At a number of places young mountain belts pass seaward into arcuate rows of islands—for example, from Honduras and Venezuela into the West Indian Islands The facts and relationships regarding island arcs that can be observed at sea differ greatly from those regarding mountains that can be observed on land Nevertheless, the observable facts attest that much the same processes of mountain building operate beneath the sea.

The island arcs are commonly bordered on their oceanward or convex sides by troughs or trenches that extend to depths of 20,000 ft or more below sea level Thus the festoons of islands are veritable mountain ranges that rise to great heights above the ocean floor The bordering trenches are presumably the oceanic counterparts of the sediment-filled geosynclines on land

On their landward or concave sides, many of the island festoons are bordered by rows of active or recently active volcanoes Great seismic activity also characterizes many of the island arcs, with shallow focus earthquakes located near the bordering trenches, intermediate depth quakes centering somewhat to the landward side of the islands, and occasional deep-focus earthquakes (at depths of from 500 to 700 km) emanating from centers still farther landward These earthquake foci thus fall into zones that slope landward beneath the islands, and these are interpreted by seismologists as shear zones along which great segments of the earth's crust are slipping past each other. This volcanic and seismic activity is evidence that island arcs are belts of active deformation of the crust—presumably orogenic processes operating today.

A notable tectonic feature of island arcs is the long, narrow, curving zone of gravity anomalies discovered in the 1920s by the Dutch geodesist, F. A. Vening Meinesz, in the East Indies and the West Indies. These zones in which the earth's gravitational pull is much less than normal follow closely along the island arcs and bordering trenches, and they indicate a deficiency of rock mass (that is, a band of abnormally lightweight rocks) below. Such a band of lighter rocks may be explained either by a tightly closed fold of moderately light crustal rocks buckled down into the heavier subcrustal rocks below (a tectogene as it has been called) or by a thick prism of very lightweight unconsolidated sediment lying below or alongside the floor of the bordering trench These two alternative explanations of the gravity anomalies imply quite different mechanisms by which the island arcs were formed.

EXPLANATIONS OF OROGENY

The empirical facts about mountains and island arcs raise many problems, and, as a result, a number of possible explanations of mountain building have been proposed.

Contraction. For nearly a century the hypothesis of a contracting earth stood essentially unchallenged as the simplest explanation of the widespread crumpling that the rocks near the earth's surface have undergone. In its traditional form this hypothesis starts with an earth that was once molten and has since cooled and shrunk so that its surface is now wrinkled like the skin of a dried apple

With increasing knowledge this classical explanation has encountered serious difficulties of a quantitative nature The amount of heat-producing radioactive materials in many common rocks is now known to be so great that nearly all if not all the heat escaping from the earth must be of radioactive origin It is no longer axiomatic, as it once seemed to be, that the earth is now and has always been cooling

Even if the presence of radioactively generated heat is ignored other serious difficulties remain For example, it may be assumed, as almost an outside estimate, that since early in its history the earth has cooled by 500C° throughout its upper 250 miles, with a volume contraction of 5%. If the circumferential shortening of about 100 miles that would result has been transmitted half way around the earth and concentrated entirely in one zone of buckling this contraction would account for the observed shortening in only one or two of the major

Distribution of mountains. Mountain chains are greatly elongated and gently to sharply curving in plan. The Western Cordillera of North America, the Andes of South America, the Alpine-Himalayan system all extend along curving trends for approximately 5000 miles before they finally pass out to sea.

Many of the higher and geologically younger mountain systems lie along the margins of continents. The Pacific Ocean is bordered by mountains—or by features characteristic of active mountain-making—almost entirely around its immense periphery. Geologically older mountain systems, now worn down to more subdued relief by long-continued stream erosion, border the Atlantic Ocean in North America and Europe.

Geosynclines. Most of the great mountain systems of the earth are located in long, narrow zones which sank and in which sedimentary rocks accumulated to great thicknesses before the mountains rose. These zones are known as geosynclines, of which the Appalachian geosyncline of eastern North America is the type example. In many geosynclines the sedimentary rocks—sandstone, mudstone, volcanic debris, limestone, and dolomite—coarsen and thicken systematically from one side to the other, indicating that these basins of deposition lay alongside uplands or mountainous areas that were undergoing stream erosion and were bordered on the opposite side by broad regions of lowland or shallow sea. As new mountains grew on the upland side, the axis of maximum down-sinking migrated gradually toward the lowland side of these basins. The widespread coincidence of early geosynclines and later mountain systems shows that the earth's crust has commonly warped slowly downward, to depths of 4, 6, or even 10 miles, before the stage of intense rock deformation that characterizes orogeny.

Crumpling of surface rocks. One of the most striking features of nearly all mountains is the extent to which the rocks exposed within them have been deformed—bent into steep-sided folds, broken by great fractures or faults, and shoved out laterally for many miles along nearly flat gliding surfaces known as overthrusts. It is as if the rocks, yielding sometimes like a brittle solid, sometimes like putty, had been crushed and squeezed between the jaws of an immense vise. If the folded and faulted strata could somehow be straightened out again into the nearly horizontal attitude they had when first deposited, they would occupy a space much wider than their present deformed width. The difference between this reconstructed original width and the present deformed width is commonly referred to as the amount of horizontal compression or crustal shortening to which the rocks have been subjected. The full significance of this compression or crustal shortening in mountainous regions is one of the great puzzles in the history of the earth.

The deformed rocks in mountainous areas exhibit not only tight folds and thrust faults, indicating such shortening, but also tensional faults and intrusions of igneous rock, indicating extension. Detailed weighing of the evidence shows that many areas have undergone compression at one time and tension at another and that some mountain ranges have rock structure that was apparently being shortened at the same time that other rocks nearby were being extended.

In many mountain systems the uppermost 1-5 miles of strata appear to have undergone much greater horizontal shortening than the basement rocks on which they rest. Thus it is by no means certain that the shortening deduced by reconstruction of the original attitude of the strata represents a true shortening of any great thickness of the earth's crust.

The exact amount of horizontal shortening or extension that the structure of any given mountain system has undergone is difficult to estimate. Partly this is because the rocks in some regions have been greatly distorted by plastic deformation and further because, even in the most rugged mountainous terrains, rock exposures are never complete in three dimensions. In spite of these sources of uncertainty there can be no question that in most mountain ranges the near-surface rocks have been greatly crumpled. Careful estimates by geologists who have

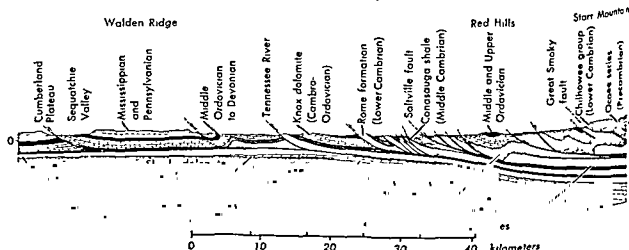


Fig. 1. Structure section across the Appalachian Mountains in Tennessee. Rock deformation largely confined to sedimentary rocks above the basement. (From P. B. King, The Tectonics of Middle North America, Princeton, 1951)

ent years in the study of a region indicate that any of the younger mountain chains have undergone a net shortening of structure, apparent or eal, of 50-100 miles or even more. It seems likely at the far more numerous older mountain systems, hich are now eroded down to their roots and thus ot accessible to similar detailed studies, have also ndergone comparable amounts of this shortening The aggregate amount of apparent structural shortening in all mountain systems, young and old, must e very large.

Rock metamorphism and igneous intrusion. In the central core areas of the more intensely deformed mountains, the minerals of which the original sedimentary rocks were composed have been metamorphosed by high temperatures and pressures and by hot solutions so that the rocks are now quartzite, slate schist, gneiss, and marble. Examples are found on the southeastern slope of the Appalachians; in the old Caledonian mountains of Scotland, Norway, and Sweden; in the southern Alps; and in the central Himalayas.

In the central core areas of other mountain systems, great masses of molten rock, or magma, have risen from deep in the crust and invaded the deformed rocks, in some places greatly altering their original texture and composition In many places these molten masses cooled underground and, having been uncovered by later stream erosion, are now represented by extensive exposures of granitic and other igneous rocks The great batholiths of the Sierra Nevada, Idaho, and British Columbia in North America and of the Andes in South America are examples. In other places, the magma broke through to the surface and formed volcanoes or poured out in great lava fields—for example, the Cascade Range and the Columbia River lava fields of northwestern United States and the great Deccan traps (plateau steps) of India.

Island arcs. The above generalizations about mountains and mountain building are based upon relations observed on land At a number of places young mountain belts pass seaward into arcuate rows of islands—for example, from Honduras and Venezuel[illegible]
and rel[illegible]
observe[illegible]
mounta[illegible]
less[illegible]
pro[illegible]
sea.

The island arcs are commonly bordered on their oceanward or convex sides by troughs or trenches that extend to depths of 20,000 ft or more below sea level Thus the festoons of islands are veritable mountain ranges that rise to great heights above the ocean floor The bordering trenches are presumably the oceanic counterparts of the sediment filled geosynclines on land.

On their landward or concave sides, many of the island festoons are bordered by rows of active or recently active volcanoes Great seismic activity also characterizes many of the island arcs, with shallow focus earthquakes located near the bordering trenches, intermediate depth quakes centering somewhat to the landward side of the islands, and occasional deep-focus earthquakes (at depths of from 500 to 700 km) emanating from centers still farther landward These earthquake foci thus fall into zones that slope landward beneath the islands, and these are interpreted by seismologists as shear zones along which great segments of the earth's crust are slipping past each other. This volcanic and seismic activity is evidence that island arcs are belts of active deformation of the crust—presumably orogenic processes operating today.

A notable tectonic feature of island arcs is the long, narrow, curving zone of gravity anomalies discovered in the 1920s by the Dutch geodesist, F. A. Vening Meinesz, in the East Indies and the West Indies These zones in which the earth's gravitational pull is much less than normal follow closely along the island arcs and bordering trenches, and they indicate a deficiency of rock mass (that is, a band of abnormally lightweight rocks) below. Such a band of lighter rocks may be explained either by a tightly closed fold of moderately light crustal rocks buckled down into the heavier subcrustal rocks below (a tectogene as it has been called) or by a thick prism of very lightweight unconsolidated sediment lying below or alongside the floor of the bordering trench. These two alternative explanations of the gravity anomalies imply quite different mechanisms by which the island arcs were formed

EXPLANATIONS OF OROGENY

The empirical facts about mountains and island arcs raise many problems, and, as a result, a number of possible explanations of mountain building have been proposed

Contraction. For nearly a century the hypothesis of a contracting earth stood essentially unchallenged as the simplest explanation of the widespread crumpling that the rocks near the earth's surface have undergone In its traditional form this hypothesis starts with an earth that was once molten and has since cooled and shrunk so that its surface is now wrinkled like the skin of a dried apple

With increasing knowledge this classical explanation has encountered serious difficulties of a quantitative nature. The amount of heat-producing radioactive materials in many common rocks is now known to be so great that nearly all if not all the heat escaping from the earth must be of radioactive origin It is no longer axiomatic, as it once seemed to be, that the earth is now and has always been cooling

Even if the presence of radioactively generated heat is ignored, other serious difficulties remain For example, it may be assumed, as almost an outside estimate that since early in its history the earth has cooled by 500C° throughout its upper 250 miles, with a volume contraction of 5%. If the circumferential shortening of about 100 miles that would result has been transmitted half way around the earth and concentrated entirely in one zone of buckling, this contraction would account for the observed shortening in only one or two of the mo[illegible]

Distribution of mountains. Mountain chains are greatly elongated and gently to sharply curving in plan. The Western Cordillera of North America, the Andes of South America, the Alpine-Himalayan system all extend along curving trends for approximately 5000 miles before they finally pass out to sea.

Many of the higher and geologically younger mountain systems lie along the margins of continents. The Pacific Ocean is bordered by mountains —or by features characteristic of active mountain-making—almost entirely around its immense periphery. Geologically older mountain systems, now worn down to more subdued relief by long-continued stream erosion, border the Atlantic Ocean in North America and Europe.

Geosynclines. Most of the great mountain systems of the earth are located in long, narrow zones which sank and in which sedimentary rocks accumulated to great thicknesses before the mountains rose. These zones are known as geosynclines, of which the Appalachian geosyncline of eastern North America is the type example. In many geosynclines the sedimentary rocks—sandstone, mudstone, volcanic debris, limestone, and dolomite—coarsen and thicken systematically from one side to the other, indicating that these basins of deposition lay alongside uplands or mountainous areas that were undergoing stream erosion and were bordered on the opposite side by broad regions of lowland or shallow sea. As new mountains grew on the upland side, the axis of maximum down-sinking migrated gradually toward the lowland side of these basins. The widespread coincidence of early geosynclines and later mountain systems shows that the earth's crust has commonly warped slowly downward, to depths of 4, 6, or even 10 miles, before the stage of intense rock deformation that characterizes orogeny

Crumpling of surface rocks. One of the most striking features of nearly all mountains is the extent to which the rocks exposed within them have been deformed—bent into steep-sided folds, broken by great fractures or faults, and shoved out laterally for many miles along nearly flat gliding surfaces known as overthrusts. It is as if the rocks, yielding sometimes like a brittle solid, sometimes like putty, had been crushed and squeezed between the jaws of an immense vise. If the folded and faulted strata could somehow be straightened out again into the nearly horizontal attitude they had when first deposited, they would occupy a space much wider than their present deformed width. The difference between this reconstructed original width and the present deformed width is commonly referred to as the amount of horizontal compression or crustal shortening to which the rocks have been subjected. The full significance of this compression or crustal shortening in mountainous regions is one of the great puzzles in the history of the earth

The deformed rocks in mountainous areas exhibit not only tight folds and thrust faults, indicating such shortening, but also tensional faults and intrusions of igneous rock, indicating extension. Detailed weighing of the evidence shows that many areas have undergone compression at one time and tension at another and that some mountain ranges have rock structure that was apparently being shortened at the same time that other rocks nearby were being extended.

In many mountain systems the uppermost 1-5 miles of strata appear to have undergone much greater horizontal shortening than the basement rocks on which they rest. Thus it is by no means certain that the shortening deduced by reconstruction of the original attitude of the strata represents a true shortening of any great thickness of the earth's crust

The exact amount of horizontal shortening or extension that the structure of any given mountain system has undergone is difficult to estimate. Partly this is because the rocks in some regions have been greatly distorted by plastic deformation and further because, even in the most rugged mountainous terrains, rock exposures are never complete in three dimensions In spite of these sources of uncertainty there can be no question that in most mountain ranges the near-surface rocks have been greatly crumpled. Careful estimates by geologists who have

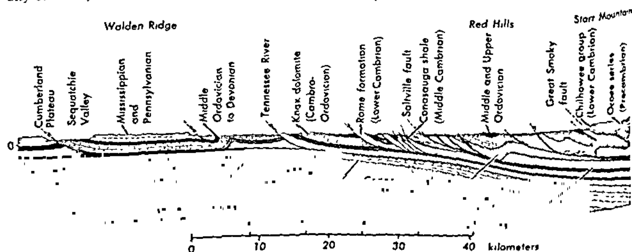


Fig 1. Structure section across the Appalachian Mountains in Tennessee Rock deformation largely confined to sedimentary rocks above the basement. (From P B King, The Tectonics of Middle North America, Princeton, 1951)

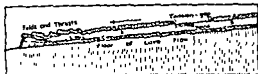


Fig 4 Section through a lava flow, showing compressional and tensional features caused by gravitational sliding (From R A. Daly, *Igneous Rocks and the Depths of the Earth*, 2d ed., McGraw-Hill, 1933)

ing crust buckles and mountain ranges form Hypotheses based on this concept bring into interesting relationship a wide range of observations. they account for the gravity anomalies along the island arcs as belts where the currents turn downward, and they are also accordant with other aspects of mountain structure and history It is fair to add, however, that the central hypothesis encounters several difficulties and that it has not yet been tested adequately. No direct physical evidence of such currents has thus far been found; and the long curving lines of mountain systems and island arcs scarcely suggest the boundaries of convection cells A laborious gathering of data on the rates of heat flow at many places—data for a map of the supposed convection cells—may be the only way the hypothesis can finally be evaluated

Gravitational sliding The idea of downhill sliding to explain many features of mountain structure has, until recently, attracted more attention in Europe than in America In brief, the concept is that thick plates of rock uplifted tectonically in a broad linear upwarp, have stretched or pulled apart under the force of gravity moved slowly down slope, and spread out laterally, somewhat like a lava flow or a giant landslide, across the adjacent lowland or geosyncline The tight folds and huge overthrusts characteristically associated with orogeny are thus attributed to secondary gravitational readjustments on the flanks of a tectonic uplift This uplift the primary cause of orogeny, may owe its origin to arching from world-wide contraction or regional compression or to local thickening of the crust caused by the rise of ... the earth's ... esis the ob... in a moun... crustal shortening but merely the aggregate horizontal distance of gravitational sliding To a greater extent than is true of some of the alternative hypotheses, the concept of gravitational sliding is subject to quantitative testing because, with sufficient attention to observable details, it should be possible to find either the gaps where the rocks may have pulled apart or else evidence that the rocks were thinned sufficiently by stretching to account for the observed horizontal displacement. Few such careful tests have yet been made.

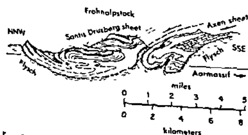


Fig 5 Thrust sheets in the Swiss Alps that have been attributed to gravitational sliding (From A Heim, *Geologie der Schweiz*, 1921, and L U De Sitter, *Structural Geology*, McGraw-Hill, 1956)

Summary. No hypothesis to explain mountain-making accounts sufficiently for all the observed facts, and for this reason, none thus far proposed has received general acceptance The underlying cause of orogeny remains one of the foremost problems of the earth sciences. *See* TECTONIC PATTERNS; TECTONOPHYSICS. [W.W.R.]

Bibliography: E. B Bailey, *Tectonic Essays, Mainly Alpine*, 1935; L. U. De Sitter, *Structural Geology*, 1956; F. J. Turner and J. Verhoogen, *Igneous and Metamorphic Petrology*, 1951; J. H. F. Umbgrove, *The Pulse of the Earth*, 1947.

Orpiment

A mineral having composition As_2S_3 and crystallizing in the monoclinic system. Crystals are small, tabular, and rarely distinct; the mineral occurs more commonly in foliated or columnar masses. There is one perfect cleavage yielding flexible folia which distinguishes it from other minerals similar in appearance The hardness is 1.5–2 (Mohs scale) and the specific gravity is 3.49. The luster is resinous and pearly on the cleavage surface; the color is lemon yellow Orpiment is associated with realgar and stibnite in veins of lead, silver, and gold ores It is found in Rumania, Peru, Japan, and Russia In the United States it occurs at Mercer, Utah, Manhattan, Nevada; and in deposits from geyser waters in Yellowstone National Park. *See* ARSENIC; REALGAR; STIBNITE. [C.S.HU.]

Orthoclase

A name for potassium feldspar ($KAlSi_3O_8$) that usually contains some sodium feldspar (up to about 50 mole % $NaAlSi_3O_8$) either in solid solution or exsolved as relatively pure $NaAlSi_3O_8$. The latter can be present as albite or analbite or in structural states intermediate between albite and analbite ... is ... acc... Th... of orthoclase may be truly monoclinic (sanidine) or can only appear to be monoclinic (normal orthoclase) according to the Al/Si distribution within the $AlSi_3O_8$ framework *See* FELDSPAR; *see also* ALBITE; PERTHITE. [F.LA.]

Orthoester

A trialkyl derivative of a nonexistent orthoacid, $RC(OH)_3$, with general formula $RC(OR)_3$.

The structure of orthocarbonates is $C(OR)_4$. Orthoesters are liquids with ethereal odors, stable to aqueous alkalies, but sensitive to acid hydrolysis. By far the most common and most important is

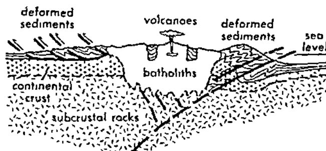


Fig. 2. Diagrammatic section through a mountain range, showing rocks crumpled and faulted by intrusions of igneous rock and by sliding down flanks of uplift formed by thickening of crust (*Modified after J. T. Wilson, An analysis of the pattern and possible cause of young mountain ranges and island arcs, Geol. Assoc. Canada Proc., 3 155, 1950*)

mountain chains of today. The shortening in these younger mountain chains has taken place in less than 2% of the time that the earth is assumed to have been cooling. The additional shortening in other mountain chains—young and old—would thus remain entirely unaccounted for. Quantitative considerations of this kind show that the contraction of a cooling earth can account for only a small part of the observed shortening, apparent or real, to which the mountain systems of the earth have been subjected.

Regional compression. Although contraction from world-wide cooling can scarcely be considered as more than a possible contributory factor in mountain-making, mountain belts may have been crumpled by regional rather than world-wide compression. The great batholiths of once-molten rock that form the cores of many mountain chains may conceivably have pushed aside the rocks they invaded and thus have caused steep folds and overthrusts in near-surface rocks that lay nearby However, some intrusions appear not to have forced their way into the overlying rocks but to have entered permissively into spaces opened for them by processes that are not well understood. Furthermore, much rock deformation, some of it very intense, occurs at great distances from any known igneous intrusions.

Whether or not igneous intrusions have pushed aside and deformed rocks near the surface, the possibility remains that redistribution of solid or molten masses deeper in the crust or even below it may cause large-scale distortion and regional compression of the surface rocks. In many seismically active areas, segments of the crust have moved laterally past one another for tens of miles or more along great breaks such as the San Andreas fault of California. Such movements demonstrate enormous shearing forces within the earth, but it is not clear how these deep-seated forces could cause the relatively shallow deformation of rocks that is characteristic of most mountain ranges.

Continental drift. The hypothesis of continental drifting attracted much interest in the 1920s but, in its original form at least, it apparently has few adherents today. By this interpretation the continents, floating like icebergs in a sea of heavier solid rock, were originally joined together as one, they have since broken up and drifted apart, crumpling up mountain ranges where they ploughed into the Pacific basin and leaving in their rear the Atlantic and Indian Oceans. The hypothesis was based upon anomalies in the past and present distribution of plant and animal species, upon extensive glaciation in the Southern Hemisphere 200 000,000 years ago, and upon the shape of the continental masses that border the Atlantic Ocean. The concept of continental drift stimulated constructive thinking and new observation, but it raised more problems than it solved It failed, for example, to account for the older mountain systems of the earth, and it depended upon physical forces which, when carefully examined, were found to be inadequate. This hypothesis has recently shown some signs of revival, although in greatly modified form, to explain new observations on the position of the earth's magnetic poles in the geologic past

Convection currents. The hypothesis that has attracted greatest attention in the past 25 years is that of convection currents in the deep interior of the earth. The postulate is simply that, given geologic time, even solid rock will turn over convectively like a boiling liquid, if it is heated more rapidly at depth than it can cool by conduction The hot rising current will spread outward at the top and drag the overlying crust along Where two convection cells meet or where for any other reason the spreading current turns downward, the overly-

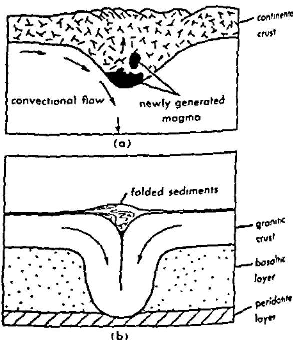
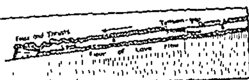


Fig 3. Sections to illustrate mountain building by convection currents. (a) D Griggs' hypothesis—down-buckling of continental crust due to convection in earth's interior (b) P. H Kuenen's and H H Hess' hypothesis—down-buckling of granitic layer (black) through basaltic layer (stippled) to impinge on peridotite shell of the deep interior (*From F. J Turner and J. Verhoogen, Igneous ar[illegible] Petrology, McGraw-Hill, 1951*)



g 4 Section through a lava flow, showing compressional and tensional features caused by gravitational iding (From R. A Daly, *Igneous Rocks and the* *epths of the Earth*, 2d ed., McGraw-Hill, 1933)

ig crust buckles and mountain ranges form Hypotheses based on this concept bring into interesting relationship a wide range of observations they ccount for the gravity anomalies along the island res as belts where the currents turn downward, nd they are also accordant with other aspects of mountain structure and history. It is fair to add, however, that the central hypothesis encounters several difficulties ...

uggest the boundaries of convection cells A laborious gathering of data on the rates of heat flow at many places—data for a map of the supposed convection cells—may be the only way the hypothesis can finally be evaluated.

Gravitational sliding. The idea of downhill sliding to explain many features of mountain structure has, until recently attracted more attention in Europe than in America In brief, the concept is that thick plates of rock, uplifted tectonically in a broad linear upwarp, have stretched or pulled apart under the force of gravity moved slowly down slope, and spread out laterally, somewhat like a lava flow or a giant landslide, across the adjacent lowland or geosyncline The tight folds and huge overthrusts characteristically associated with orogeny are thus attributed to secondary gravitational readjustments on the flanks of a tectonic uplift This uplift, the primary cause of orogeny, may owe its origin to arching from world wide contraction or regional compression or to local thickening of the crust caused by the ... the earth ... esis the ... in a mo ... crustal shortening but merely the aggregate horizontal distance of gravitational sliding. To a greater extent than is true of some of the alternative hypotheses, the concept of gravitational sliding is subject to quantitative testing because, with sufficient attention to observable details, it should be possible to find either the gaps where the rocks may have pulled apart or else evidence that the rocks were thinned sufficiently by stretching to account for the observed horizontal displacement Few such careful tests have yet been made.

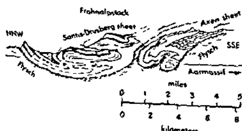


Fig 5 Thrust sheets in the Swiss Alps that have been attributed to gravitational sliding. (From A. Heim, *Geologie der Schweiz*, 1921; and L U De Sitter, *Structural Geology*, McGraw-Hill, 1956)

Summary. No hypothesis to explain mountain-making accounts sufficiently for all the observed facts, and for this reason, none thus far proposed has received general acceptance. The underlying cause of orogeny remains one of the foremost problems of the earth sciences See TECTONIC PATTERNS; TECTONOPHYSICS. [W.N.R.]

Bibliography: E. B. Bailey, *Tectonic Essays, Mainly Alpine*, 1935, L. U De Sitter, *Structural Geology*, 1956; F. J Turner and J. Verhoogen, *Igneous and Metamorphic Petrology*, 1951; J. H. F Umbgrove, *The Pulse of the Earth*, 1947.

Orpiment

A mineral having composition As_2S_3 and crystallizing in the monoclinic system Crystals are small tabular, and rarely distinct, the mineral occurs more commonly in foliated or columnar masses. There is one perfect cleavage yielding flexible folia which distinguishes it from other minerals similar in appearance The hardness is 1.5–2 (Mohs scale) and the specific gravity is 3 49 The luster is resinous and pearly on the cleavage surface, the color is lemon yellow Orpiment is associated with realgar and stibnite in veins of lead, silver, and gold ores. It is found in Rumania, Peru, Japan, and Russia In the United States it occurs at Mercer, Utah, Manhattan, Nevada, and in deposits from geyser waters in Yellowstone National Park. See ARSENIC, REALGAR; STIBNITE [C.S.HU.]

Orthoclase

A name for potassium feldspar ($KAlSi_3O_8$) that usually contains some sodium feldspar (up to about 50 mole % $NaAlSi_3O_8$) either in solid solution or exsolved as relatively pure $NaAlSi_3O_8$ The latter can be present as albite or analbite or in structural states intermediate between albite and analbite. If exsolution is detectable, such material is called cryptoperthite, microperthite, or perthite according to increasing size of the exsolved areas The symmetry of orthoclase may be truly monoclinic (sanidine) or can only appear to be monoclinic (normal orthoclase) according to the Al/Si distribution within the $AlSi_3O_8$ framework See FELDSPAR, see also ALBITE; PERTHITE. [F.LA.]

Orthoester

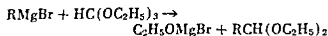
A trialkyl derivative of a nonexistent orthoacid, $RC(OH)_3$, with general formula $RC(OR)_3$.

The structure of orthocarbonates is $C(OR)_4$. Orthoesters are liquids with ethereal odors, stable to aqueous alkalies, but sensitive to acid hydrolysis By far the most common and most important is

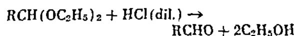
ethyl orthoformate, $HC(OC_2H_5)_3$, although ethyl orthoacetate and higher members are known.

Orthoformate esters are manufactured by interaction of chloroform with an excess of the requisite sodium alkoxide; thus, ethyl orthoformate results from the reaction of chloroform and sodium ethoxide. Higher orthoesters are usually made by treatment of imidoester hydrochlorides (from nitriles, hydrogen chloride, and alcohol) with excess alcohol. Orthocarbonates cannot be made from carbon tetrachloride; instead, chloropicrin, CCl_3NO_2, is heated with the requisite sodium alkoxide.

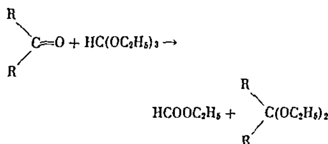
Orthoformate esters are frequently used in synthetic organic chemistry, for example, in preparing aldehydes from Grignard reagents (*see* GRIGNARD REACTION).

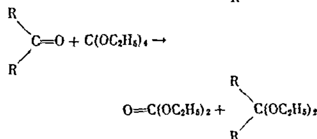


Then,

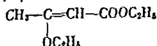


To make ketals from ketones, either an orthoformate or an orthocarbonate can be used,





Enol ethers of β-keto esters are readily formed from ethyl orthoformate and the β-keto ester, for example, ethyl β-ethoxycrotonate,



from ethyl acetoacetate. *See* ETHYL ACETOACETATE

Ethyl orthocarbonate, $C(OC_2H_5)_4$, finds use in the preparation of higher orthoesters by reaction with Grignard reagents *See* ESTER [F.B.R.]

Orthogonal polynomials

A special case of orthogonal functions that arise in many physical problems (often as the solutions of differential equations), in the study of distribution functions, and in certain other situations where one approximates fairly general functions by polynomials. *See* PROBABILITY.

Each set of orthogonal polynomials is defined with respect to a particular averaging procedure. The average value of a suitable function f is denoted by $E\{f\}$. Examples are

$$E\{f\} = \frac{1}{2}\int_{-1}^{1} f(x)\,dx \qquad (1)$$

$$E\{f\} = \frac{\int_{-1}^{1} f(x)(1-x)^{\alpha}(1+x)^{\beta}\,dx}{\int_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\,dx} \qquad (\alpha,\beta > -1) \qquad (2)$$

$$E\{f\} = \int_{0}^{\infty} f(x)e^{-x}\,dx \qquad (3)$$

$$E\{f\} = (2\pi)^{-1/2}\int_{-\infty}^{\infty} f(x)e^{-x^2}\,dx \qquad (4)$$

In general an averaging procedure has the form

$$E\{f\} = \int_{-\infty}^{\infty} f(x)\,d\sigma(x)$$

a Stieltjes integral, where σ is a distribution function, that is, an increasing function with $\sigma(-\infty) = 0$ and $\sigma(+\infty) = 1$. In the above examples σ has the form $\sigma(x) = b^{-1}\int_{-\infty}^{x}\omega(y)\,dy$ where ω is a nonnegative weight function and $b = \int_{-\infty}^{\infty}\omega(y)\,dy$. Consideration will be given only to averaging procedures for which all the moments $\mu_n = E\{x^n\} = \int_{-\infty}^{\infty} x^n\,d\sigma(x)$ exist and for which $E\{|P|\} > 0$ for every polynomial P.

Orthogonal functions. Two functions f and g are said to be orthogonal with respect to a given averaging procedure if $E\{f\bar{g}\} = 0$ where the bar denotes complex conjugation. By the system of orthogonal polynomials associated with the averaging procedure is meant a sequence $P_0, P_1, P_2, \ldots$ of polynomials P_n having exact degree n, which are mutually orthogonal, that is, $E\{P_m\bar{P}_n\} = 0$ for $m \neq n$. This last condition is equivalent to the statement that each P_n is orthogonal to all polynomials of degree less than n. Thus P_n has the form $P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ where $a_n \neq 0$ and is subject to the n conditions $E\{x^kP_n\} = 0$ for $k = 0, 1, \ldots, n-1$. This gives n linear equations in the $n+1$ coefficients of P_n, leaving one more condition, called a normalization, to be imposed. The method of normalization differs in different references. Orthogonal polynomials arising from the average of the form (1), Legendre polynomials, satisfy Legendre's differential equation. With the normalization $P_n(1) = 1$ the first few Legendre polynomials are $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$, $P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x$. The average in Eq. (1) is the special case of Eq. (2) with $\alpha = \beta = 0$, the orthogonal polynomials corresponding to averages of the form (2) are called Jacobi polynomials, those associated with (3), Laguerre polynomials; with (4), Hermite polynomials.

The proper setting for the study of expansions in terms of orthogonal polynomials is the Hilbert space H of functions f such that $E\{|f|^2\}$ exists and is finite. The inner product is $(f,g) = E\{f\bar{g}\}$. In analogy with the procedure for Fourier series one can write down a formal expansion

$$f(x) \sim \sum \qquad (5)$$

where the coefficients are given by the formula

$$c_n = E\{f\bar{P}_n\}/E\{|P_n|^2\} \qquad (6)$$

The Nth partial sum of the series (5),

$$s_N(x) = \sum_0^N c_n P_n(x)$$

has the property that among all polynomials p of degree not exceeding N, the minimum of the quadratic deviation $E\{|f - p|^2\}$ is achieved uniquely by $p = s_N$. If the only function f in H with the property that $E\{x^k f\} = 0$ for every k is the zero function, one says that the polynomials are "complete" in H. In this case the coefficients in (6) uniquely determine the function f, and the properties of the series (5) are quite analogous to the properties of Fourier series of functions in L^2 (see FOURIER SERIES AND INTEGRALS) The polynomials are always complete when the average is taken over a finite interval, but in general some extra assumption is required. The divergence of the series $\Sigma\mu_{2n}^{-1/2n}$ is a sufficient condition for the completeness of the polynomials (It is fulfilled in each of the examples cited)

The orthogonality property entails certain algebraic properties for the polynomials. For example, the zeros of P_n are all distinct, they lie in the interior of the interval over which the average is taken, and they separate the zeros of P_{n-1} Let $X_1^{(n)}, \ldots, X_n^{(n)}$ be the zeros of P_n One can find constants $b_1^{(n)}, \ldots, b_n^{(n)}$ such that $Q_n\{1\} = 1, Q_n\{P_k\} = 0$ for $0 < k < n$, where $Q_n\{f\} = \sum_{j=1}^{n} b_j^{(n)} f(X_j^{(n)})$ In the case of an average over a finite interval, $\lim_{n\to\infty} Q_n\{f\} = E\{f\}$ for every continuous f This is of interest in approximate integration, because the integral $E\{f\}$ is approximated by an expression $Q_n\{f\}$ which depends only on the values of f at n points and, what is remarkable, $Q_n\{f\} = E\{f\}$ whenever f is a polynomial of degree $\leq 2n - 1$ whereas one would ordinarily expect an n-point approximation to be exact only for polynomials of degree $\leq n$

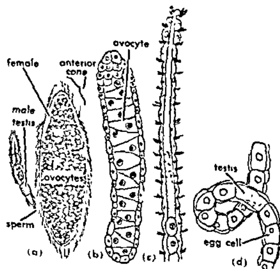
Ultraspherical polynomials. There exists a theory of orthogonal polynomials in several variables The most important applications involve averages over spheres in m dimensions Complete sets of orthogonal polynomials may be chosen among the homogeneous, harmonic polynomials. A polynomial $P(x_1, \ldots, x_m)$ is homogeneous of degree n if $P(\lambda x_1, \ldots, \lambda x_m) = \lambda^n P(x_1, \ldots, x_m)$ for each λ, it is harmonic if it satisfies Laplace's differential equation Let P be such a polynomial with the property that $P(1, 0, \ldots, 0) \neq 0$ Consider $P(x, y_1, \ldots, y_{m-1})$ as a polynomial in the $m - 1$ variables $y_1, \ldots, y_{m-1}$ and take the average over a sphere centered at the origin in $m - 1$ dimensions. The result is a polynomial $P_n(x)$ of degree n The orthogonality over the sphere in m dimensions translates itself into orthogonality on the interval $[-1,1]$ with the weight function $\omega(x) = (1 - x^2)^{(m-3)/2}$. For fixed m, the polynomials obtained this way are the ultraspherical polynomials, special cases of the Jacobi polynomials with $\alpha = \beta = (m - 3)/2$. Where $m = 3$ corresponds to 3-dimensional space, these are Legendre polynomials. *See* DIFFERENTIAL EQUATION; GEOMETRY, RIEMANNIAN; LAPLACE'S DIFFERENTIAL EQUATION; POLYNOMIAL SYSTEMS OF EQUATIONS [C.S H.]

Bibliography: J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, Math. Survs 1, 1943; G. Szegö, *Orthogonal Polynomials*, A.M.S. Colloqu. Publ 23, 1939.

Orthonectida

An order of Mesozoa The orthonectids parasitize various marine invertebrates as multinucleate plasmodia, sometimes causing considerable damage to the host The plasmodia multiply by fragmentation. Eventually they give rise asexually, by polyembryony, to sexual males and females. Commonly only one sex arises from a given plasmodium.

These sexually mature forms escape as minute ciliated organisms. Structurally, they are composed of a single layer of ciliated epithelial cells surrounding an inner mass of sex cells The ciliated cells are disposed in rings around the body Those at the anterior end form the anterior cone, on which the cilia are directed forward, while on the rest of the body they point backward Males are smaller than the corresponding females. A few species, however, are hermaphroditic.



Orthonectids (a) Typical orthonectid, *Rhopalura ophiocomae*, male discharging sperm near the genital pore of the female (b) *R. metschnikovi*, female, germ cells oriented in a double row (c) *Stoechartrum giardia*, anterior end (d) *S. giardia*, part of trunk showing egg cells in single linear series, one to each apparent segment In one of the segments a testis has developed

After insemination, the eggs develop in the female and form ciliated larvae. When liberated, these larvae invade new individuals of their host and then disaggregate, liberating germinal cells which give rise to new plasmodia *See* MESOZOA; *see also* REPRODUCTION, ANIMAL. [B.H M.]

Orthoptera

An order of generalized pterygote, winged insects, characterized by gradual metamorphosis, chewing mouthparts, and two pairs of wings. The front wings are thickened and leathery while the second pair are thin, membranous, and folded fanwise beneath the first. Many forms have reduced wings, and in others the wings are entirely absent. Most Orthoptera are plant-eaters. Although some may inhabit marshy areas, none is aquatic in the true sense.

Orthoptera are among the noisiest of insects, possessing both well-developed sound-producing and sound-perceiving organs. The Locustidae produce sound by rubbing a series of tiny pegs on the inner faces of the femora against the outer wings. The auditory organs are on either side of the base of the abdomen. In most of the remaining Orthoptera sound is produced by rapidly rubbing the wings together, and the auditory organs are located at the bases of the front tibiae.

Habitats of the Orthoptera are variable Representatives of most families are found on the surface of the ground However, the Grylloblattidae occur under stones or in loose soil, the Tridactilidae in sand or mud, and the Mantidae in grasses or herbaceous cover. Members of these families are not of great economic importance; families which cause considerable economic damage to man are given in the table.

Habitats of some Orthoptera of economic importance

Name	Family habitats and damage by species
Locustidae	Live in grasses, herbaceous cover, or trees
Red-legged grasshopper, *Melanoplus femur-rubrum*	North America, feeds on grasses and many cultivated crops
Migratory grasshopper, *Melanoplus mexicanus*	N A ; feeds on grasses and many cultivated crops
High plains grasshopper, *Dissosleira longipennis*	N A ; destroys range in high plains just east of Rocky Mts
Clear-winged grass-hopper, *Camnula pellucida*	N A ; most abundant in Northwest at higher elevations, destroys range and cultivated crops
Tettigoniidae	Live in caves, under stones, in loose soil, grasses, herbaceous cover, or trees
Mormon cricket, *Anabrus simplex*	N A , most abundant north and west of Utah and Colorado into Canada; attacks most field and garden crops
Gryllidae	Live under stones, in logs, ant nests, grasses, herbaceous cover, trees, or human dwellings
Field cricket, *Achela assimilis*	North and South America, may attack most crops, at times it enters houses where it may eat holes in cloth or paper
Snowy tree cricket, *Oecanthus niveus*	N A ; weaken twigs and canes of various fruits by egg laying in the tissues
Gryllotalpidae	Live in sand or mud
Northern mole cricket, *Gryllotalpa hexadactyla*	N.A.; disturbs seedlings and eats their roots in moist light soils
Phasmidae	Live in trees
Walkingstick, *Diapheromera femorala*	N A , may become sufficiently abundant to defoliate trees
Blattidae	Live under stones, in logs, or in human dwellings
German cockroach, *Blattella germanica*	Cosmopolitan, live in human dwellings where they feed on various food products, paper book bindings, etc , and contaminate food
Oriental cockroach, *Blatta orientalis*	Cosmopolitan; damage similar to *B germanica*
American cockroach, *Periplaneta americana*	Cosmopolitan, damage similar to *B germanica*
Brown-banded roach, *Supella supellectilium*	A more recently introduced species, damage similar to that of other roaches but habits differ—it prefers upper part of house to damp basement and water pipes

FAMILIES OF ORTHOPTERA

The number of families included in this order varies because the species differ so in shape and form. Thus, several families have been given ordinal status by some authorities.

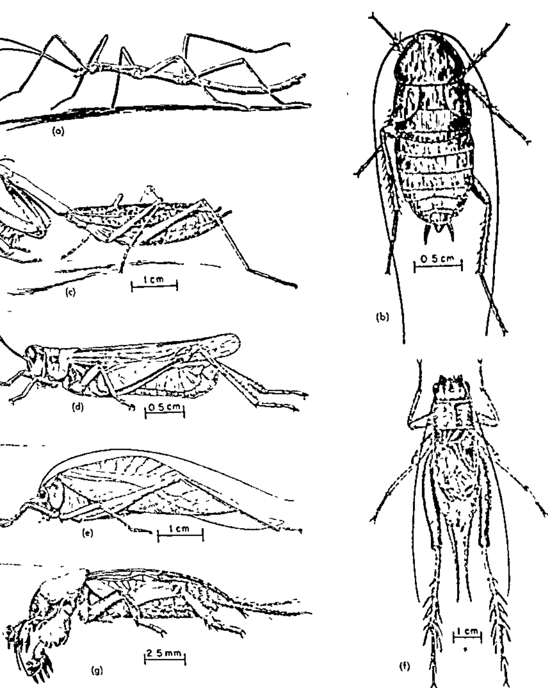
Phasmidae. The walkingsticks and leaf insects are examples of this family Many are remarkable mimics of dead twigs, being long, slender, wingless, and brownish in color. Occasionally they become so abundant that they defoliate trees Some tropical forms are winged, the wings bearing a striking resemblance to leaves.

Blattidae. This family includes the cockroaches and are flattened insects with a large, scalelike pronotum which overhangs the head. Some cockroaches are wingless, but most of them are winged A few species live in human dwellings and are among the most abundant and persistent of household pests. Several eggs are usually laid in a capsule, but some species are ovoviviparous, the eggs hatching within the body of the female

Mantidae. This family includes the praying mantids, characterized by a long slender prothorax bearing a pair of large grasping legs, and a freely moving head with large eyes. Mantids are predaceous, clutching the prey with the large forelegs and bringing it to the mouth. Eggs are laid in masses called oothecae, which are attached to twigs.

Grylloblattidae. A single genus, *Grylloblatta*, constitutes this family. They are found beneath rocks and in similar situations in mountains of western North America and Japan Grylloblattids are slender, wingless, about an inch in length, and as in the preceding families, have hindlegs not developed for jumping. They are strongly cold adapted, *G campodeiformis* seeking a temperature of approximately 4°C when given a choice They are found at the ground surface in the autumn and sometimes in the spring During the summer they apparently descend into the soil or rock slides.

Locustidae. The true grasshoppers and the migratory locusts are members of this family. The



Examples of Orthoptera (a) A walkingstick, *Diapheromera femorata* (b) A cockroach, *Blatta orientalis* (c) A praying mantid, *Stagomantis carolina* (d) A grasshopper, *Melanoplus mexicanus* (e) A meadow grasshopper, *Microcentrum rhombifolium*. (f) A grass cricket, *Nemobius fasciatus* (g) A mole cricket, *Gryllotalpa hexadactyla* (*From Illinois Natural History Survey*)

antennae are usually less than half of the length of the body The hindlegs are adapted for jumping, and the ovipositor is a multipartite structure with which the female works her abdomen into the ground The female gradually withdraws it as she lays eggs in a subterranean "pod" To this group belong the majority of crop-destroying Orthoptera In North America the most injurious species belong to the genus *Melanoplus*. In Africa and Asia devastating species belong to the genus *Schistocerca* Some of these species migrate long distances

Tetrigidae. These are tiny grasshoppers, called grouse locusts or pygmy grasshoppers, which superficially resemble the Locustidae The front wings, however, are reduced to small scalelike structures, and the pronotum extends backward over the whole body

Tettigoniidae This family consists of the longhorn or green grasshoppers The antennae are usu-

ally as long as or longer than the body, and most of the common species are green in color. The hindlegs are fitted for jumping; the ovipositor is elongate and vertically flattened. While most of the Locustidae are diurnal, most of the Tettigoniidae are active at night. To this group belong the katydids, meadow grasshoppers, cave or camel crickets, and the Mormon cricket. This latter species, which occurs in the Rocky Mountain area, is occasionally very destructive to crops.

Gryllidae. The true crickets, with which most people are familiar, belong to this family. These are usually rather chunky, dark-colored insects with long antennae and long cylindrical ovipositors. Field crickets of the genus *Acheta* are common in many habitats and may enter houses. Eggs are laid singly in the soil. Tree crickets of the genus *Oecanthus live above the ground They depart from* the darker coloration of many of their relatives and are usually pale green. With their cylindrical ovipositors they pierce the stems of bushes or trees, where their eggs are laid.

Gryllotalpidae. The mole crickets, insects highly specialized for fossorial existence, are in this family. They burrow just beneath the soil, pushing up long ridges as do moles.

Tridactylidae. These are the pigmy mole crickets, *which have similar habits, but they are about a* fifth the length of the Gryllotalpidae, and usually are found in moist soil near water. [H.B.MI.]

Orthoquartzite

A rock (also known as quartzose sandstone) composed almost entirely of detrital quartz grains, it is generally considered that over 95% of the detrital grains must be quartz for it to fall in this group. Orthoquartzite is a sedimentary rock distinct from *metamorphic quartzite (metaquartzite, a metamorphic* rock formed at high temperature or pressure).

Mineral composition. Any clay that is present in orthoquartzites is insignificant in amount Feldspar is either absent or present in very small amounts. Other unstable mineral grains are not common. The minor accessory minerals are only the most stable ones, zircon and tourmaline The texture of orthoquartzites, aside from the lack of clay matrix, tends to be distinctive. Typically, the grains are extremely well rounded and tend to have high sphericity as well. The sorting of the various sizes is very good Some of the best sorted sands that have been described belong in this category. Any rock fragments are almost invariably chert. [illegible] dominantly cross-bedded [illegible]

that is, without many bedding planes. The local variation in mineral composition and texture laterally (along beds) and vertically (from bed to bed) is small Regionally, orthoquartzite *formations tend* to be homogeneous, showing only small changes in heavy mineral suites that may be due to contributions from different source areas.

Mineral cements. The particles of orthoquartzites are bound together by precipitated mineral cements. The most abundant mineral cement is quartz. Normally quartz cement occurs as secondary overgrowths on detrital grains, the overgrowth being deposited as a single crystal, optically continuous with the detrital grain crystal. The overgrowths may show euhedral crystal faces if they have grown into open pore spaces. Those sandstones that are so completely cemented by secondary quartz that practically no pores remain show boundaries between adjacent overgrowths that are irregular as the result of interference during crystal growth. In some orthoquartzites the silica cement is chert or opal. Not all orthoquartzites are well cemented; many are friable and almost unindurated.

Carbonate minerals are important as cementing agents in orthoquartzites. Calcite and dolomite are *most abundant but siderite*, iron-rich dolomite, and aragonite may also be present. Carbonate cement frequently has replaced some detrital quartz; that is, some of the original detrital quartz has been dissolved and carbonate precipitated in its place In some orthoquartzites the carbonate cement is so abundant that quartz grains do not touch each other and appear to float in the mineral cement. This may happen as a result of extensive replacement of detrital quartz or as a result of original sedimentation *conditions under which* a mixture of quartz grains and carbonate grains settled together, the carbonate grains later having lost their identity by recrystallization.

Occurrence. Typically, orthoquartzites are associated with fossiliferous limestone and calcareous shale beds. Orthoquartzites are most abundant in thin sedimentary sections, primarily in the interior of continents, but may also be found as the early deposits in some geosynclines They range in age *from Precambrian to Tertiary* but appear to be more abundant in the Precambrian and Paleozoic than in the Mesozoic and Tertiary.

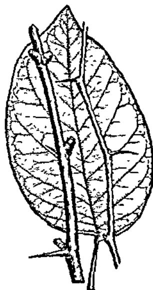
Origin. The chief inference from the mineralogical composition (all quartz) and texture (well-sorted, well-rounded) is that the orthoquartzites represent material that has been subjected to a great deal of chemical weathering at the source and to abrasion and sorting during transportation and deposition One way by which material of this composition and texture can be produced is by the slow erosion of a low-lying, tectonically stable source area. Under these conditions chemical weathering has had sufficient time to eliminate all unstable minerals, and extensive abrasion and sorting by shoreline processes would give a well sorted well-rounded sandstone. A second way in which orthoquartzites can be produced is by repeated reworking of older sediments As the sandstones go through several cycles of erosion, transportation, and deposition, they become progressively better sorted, and most unstable minerals are lost. Second cycle quartz grains, grains that show abraded secondary quartz overgrowths, are evidence of derivation from preexisting sediments *See* ARKOSE; GRAYWACKE; SANDSTONE; [illegible]ARY ROCKS; SUBGRAYWACKE. [R.S.]

[Or]thorhombic pyroxene

[Th]e general name for the solid solution series be[tw]een the end members enstatite, $MgSiO_3$, and fer[ros]ilite, $FeSiO_3$, also known as orthopyroxene or [ens]tatite. The term hypersthene is often used to [ind]icate the intermediate compositions. Oriented [inc]lusions occasionally produce a bronze luster in [the] crystals; the mineral is then called bronzite [Th]e orthorhombic pyroxenes form prismatic crys[tal]s that are green, gray-green, brown, or black in [col]or, with the 87° pyroxene (110) cleavages Thin [sec]tions of the mineral are often colorless or may [sho]w a weak pleochroism (color change on rota[tio]n in plane polarized light). The parallel extinc[tio]n distinguishes the mineral from the monoclinic [py]roxenes Oriented inclusions and exsolution la[me]llae of diopsidic or augitic materials are usually [pr]esent See PYROXENE, see also DIOPSIDE, ENSTA[TIT]E; PIGEONITE [G W D]

[O]sage-orange

[A] genus, *Maclura*, of the mulberry family, with one [sp]ecies, *M pomifera*. This tree may attain a height [of] 60 ft and has yellowish bark, milky sap, simple [ent]ire leaves, strong axillary thorns, and aggregate [gr]een fruit about the size and shape of an orange [It] is planted for hedges and as an ornamental, es[p]ecially in the eastern United States where it is [n]aturalized The wood is used for fence posts and [f]uel and as a source of a yellow dye. It has also [b]een used for archery bows, hence one of its com[m]on names, bow wood See FOREST AND FORESTRY; TREE [A H G.]



Osage-orange, *Maclura pomifera*. (A H Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

Oscillation

Any effect that varies in a back-and-forth or reciprocating manner. Examples of oscillation include the variation[s] ... re in a sound wave and the fluctuations in a mathematical function whose value repeatedly alternates above and below some mean value.

The term oscillation is for most purposes synonymous with vibration, although the latter sometimes implies primarily a mechanical motion. A device designed to reduce a person's weight by shaking him is likely to be called a *vibrator*, whereas an electronic device that produces an electric current which reverses its direction periodically is usually called an oscillator. The *alternating current* and the associated electric and magnetic fields are referred to as electric (or electromagnetic) oscillations.

If a system is set into oscillation by some initial disturbance and then left alone, the effect is called a free oscillation. A *forced oscillation* is one in which the oscillation is in response to a steadily applied periodic disturbance.

Any oscillation that continually decreases in amplitude, usually because the oscillating system is sending out energy, is spoken of as a damped oscillation An oscillation that maintains a steady amplitude, usually because of an outside source of energy, is undamped See ANHARMONIC OSCILLATOR; DAMPING, FORCED OSCILLATION; HARMONIC OSCILLATOR; MECHANICAL VIBRATION; OSCILLATOR; VIBRATION. [J M KE.]

Oscillator

An electronic circuit that converts energy from a direct-current source into a periodically varying electrical output. If the output voltage is a sine-wave function of time, the generator is called a sinusoidal, or harmonic, oscillator. Only sinusoidal oscillators are discussed in this article. If the output waveform contains abrupt changes in voltage, such as occur in a pulse or square wave, the device is called a relaxation oscillator. See RELAXATION OSCILLATOR, WAVE-SHAPING CIRCUITS.

Basic principles. The fundamental laws governing sinusoidal oscillators are the same for all oscillator circuits. These basic concepts are illustrated in Fig 1 The amplifier provides an output voltage e_o as a consequence of an external input signal voltage e_s. The voltage e_o is applied to a circuit called a feedback network whose output is e_f If the feedback voltage e_f were made identically equal to the input voltage e_s, and if the external input were disconnected and the feedback voltage connected to the amplifier input terminals 1 and 2, the amplifier would continue to provide the same output voltage e_o as before This requires that the instantaneous values of e_f and e_s be exactly equal at all times. Since no restriction was made on the waveform, it need not be sinusoidal.

If the entire circuit operates linearly and the amplifier or feedback network or both contain reactive elements, the only periodic wave that will preserve its form is the sinusoidal waveform, and such a circuit will be a sinusoidal oscillator For sinusoidal oscillators the condition where e_s equals e_f requires that amplitude, phase, and frequency of e_s and e_f be identical The phase shift ...

ally as long as or longer than the body, and most of the common species are green in color. The hindlegs are fitted for jumping; the ovipositor is elongate and vertically flattened. While most of the Locustidae are diurnal, most of the Tettigoniidae are active at night. To this group belong the katydids, meadow grasshoppers, cave or camel crickets, and the Mormon cricket. This latter species, which occurs in the Rocky Mountain area, is occasionally very destructive to crops

Gryllidae. The true crickets, with which most people are familiar, belong to this family. These are usually rather chunky, dark-colored insects with long antennae and long cylindrical ovipositors. Field crickets of the genus *Acheta* are common in many habitats and may enter houses. Eggs are laid singly in the soil. Tree crickets of the genus *Oecanthus* live above the ground They depart from the darker coloration of many of their relatives and are usually pale green. With their cylindrical ovipositors they pierce the stems of bushes or trees, where their eggs are laid.

Gryllotalpidae. The mole crickets, insects highly specialized for fossorial existence, are in this family. They burrow just beneath the soil, pushing up long ridges as do moles.

Tridactylidae. These are the pigmy mole crickets, which have similar habits, but they are about a fifth the length of the Gryllotalpidae, and usually are found in moist soil near water. [H.B MI]

Orthoquartzite

A rock (also known as quartzose sandstone) composed almost entirely of detrital quartz grains; it is generally considered that over 95% of the detrital grains must be quartz for it to fall in this group. Orthoquartzite is a sedimentary rock distinct from metamorphic quartzite (metaquartzite, a metamorphic rock formed at high temperature or pressure).

Mineral composition. Any clay that is present in orthoquartzites is insignificant in amount. Feldspar is either absent or present in very small amounts. Other unstable mineral grains are not common. The minor accessory minerals are only the most stable ones, zircon and tourmaline. The texture of orthoquartzites, aside from the lack of clay matrix, tends to be distinctive. Typically, the grains are extremely well rounded and tend to have high sphericity as well. The sorting of the various sizes is very good. Some of the best sorted sands that have been described belong in this category. Any rock fragments are almost invariably chert. [illegible]

variation in mineral composition and texture laterally (along beds) and vertically (from bed to bed) is small. Regionally, orthoquartzite formations tend to be homogeneous, showing only small changes in heavy mineral suites that may be due to contributions from different source areas.

Mineral cements. The particles of orthoquartzites are bound together by precipitated mineral cements. The most abundant mineral cement is quartz. Normally quartz cement occurs as secondary overgrowths on detrital grains, the overgrowth being deposited as a single crystal, optically continuous with the detrital grain crystal. The overgrowths may show euhedral crystal faces if they have grown into open pore spaces. Those sandstones that are so completely cemented by secondary quartz that practically no pores remain show boundaries between adjacent overgrowths that are irregular as the result of interference during crystal growth. In some orthoquartzites the silica cement is chert or opal. Not all orthoquartzites are well cemented; many are friable and almost unindurated.

Carbonate minerals are important as cementing agents in orthoquartzites. Calcite and dolomite are most abundant but siderite, iron-rich dolomite, and aragonite may also be present. Carbonate cement frequently has replaced some detrital quartz; that is, some of the original detrital quartz has been dissolved and carbonate precipitated in its place. In some orthoquartzites the carbonate cement is so abundant that quartz grains do not touch each other and appear to float in the mineral cement. This may happen as a result of extensive replacement of detrital quartz or as a result of original sedimentation conditions under which a mixture of quartz grains and carbonate grains settled together, the carbonate grains later having lost their identity by recrystallization.

Occurrence. Typically, orthoquartzites are associated with fossiliferous limestone and calcareous shale beds Orthoquartzites are most abundant in thin sedimentary sections, primarily in the interior of continents, but may also be found as the early deposits in some geosynclines. They range in age from Precambrian to Tertiary but appear to be more abundant in the Precambrian and Paleozoic than in the Mesozoic and Tertiary.

Origin. The chief inference from the mineralogical composition (all quartz) and texture (well sorted, well-rounded) is that the orthoquartzites represent material that has been subjected to a great deal of chemical weathering at the source and to abrasion and sorting during transportation and deposition One way by which material of this composition and texture can be produced is by the slow erosion of a low-lying, tectonically stable source area. Under these conditions chemical weathering has had sufficient time to eliminate all unstable minerals, and extensive abrasion and sorting by shoreline processes would give a well-sorted, well-rounded sandstone A second way in which orthoquartzites can be produced is by repeated reworking of older sediments. As the sandstones go through several cycles of erosion, transportation and deposition, they become progressively better sorted, and most unstable minerals are lost Second-cycle quartz grains, grains that show abraded secondary quartz overgrowths, are evidence of derivation from preexisting sediments. *See* ARKOSE; GRAYWACKE; SANDSTONE; SEDIMENTARY ROCKS; SUBGRAYWACKE. [illegible]

In fact, at the higher frequencies, the impedance of the phase-shifting network may become quite small, and the loading of the amplifier by the phase-shifting network may become serious. On the other hand, if R and C are made large, but still well within the range of commercially available values, frequencies of one or two cycles are easily attained. Inductors suitable for use in LC tuned oscillators for this frequency range are often impractical.

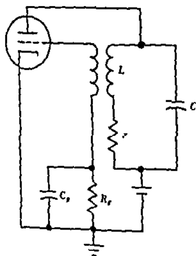


Fig 3 A tuned plate oscillator.

The frequency of the oscillator may be varied by changing the value of any of the impedance elements in the phase-shifting network. But if circuit components are varied without discrimination, the impedance looking into the phase shifting network and the magnitude of the transfer function will change, and there is the possibility that βA will fall below unity and the circuit will stop oscillating. On the other hand, if βA should continue to increase beyond unity, the excursion of tube voltages must be farther into the range of nonlinear operation to limit the amplitude of oscillation, as a result, the amplitude must increase. Hence, a change in frequency occasioned by an arbitrary variation of circuit parameters will usually affect the amplitude. For small variations of frequency a variation of any single circuit component is quite feasible. But for variations of frequency over a large range the three capacitors must be varied simultaneously. Such a variation will keep the input impedance to the phase-shifting network constant and keep constant also the magnitude of β A variation of all three resistors simultaneously will keep β constant, but the impedance will vary. Such a variation of the impedance will vary A and consequently βA. The attenuation of the phase-shifting network can be reduced by using more than three sections in the phase-shifting network or by removing the restrictions that all the capacitors be equal and that all the resistors be equal, but each of these methods complicates the matter of obtaining variable-frequency operation.

The phase-shift oscillator is usually operated in Class A to keep distortion to a minimum (*see* AMPLIFIER). Self-bias is obtained from the cathode $R_k - C_k$ combination in Fig. 2. *See* BIAS (ELECTRON TUBE).

Resonant-circuit oscillators. Figure 3 shows the tuned plate oscillator, in which a resonant circuit is used to determine the frequency. In Fig. 3 r represents a resistance in series with the plate winding (of inductance L) to account for the losses in the transformer. If these losses are negligible so that r can be neglected, then at the frequency $\omega = 1/\sqrt{LC}$ the impedance of the resonant circuit is arbitrarily large and purely resistive. *See* RESONANCE (ALTERNATING-CURRENT CIRCUITS). The voltage drop across the inductor from plate to ground is precisely 180° out of phase with the applied input voltage to the vacuum tube. If the secondary winding of the transformer is connected to introduce an additional phase shift of 180° (it is assumed that the secondary is not loaded), the total loop phase shift is exactly zero. At this frequency, then, the phase-shift condition for oscillation will have been satisfied. Since the transformer is considered to be unloaded, the ratio of the amplitude of the secondary to the primary voltage is M/L, where M is the mutual inductance. Since $A = \mu$ for an amplifier with an infinite load impedance (the resonant condition), βA equals 1 if L/M is made equal to μ.

The above considerations emphasize that the criteria stated with respect to the loop phase shift and the loop gain are the conditions which characterize the operation of the circuit. In particular, the frequency of oscillation is in the neighborhood of, but in no way simply related to, the frequency of a "natural" oscillation that might be excited in the resonant circuit. Neither is there any a priori connection between the oscillation frequency and the steady-state resonance frequency. The frequency of oscillation is determined solely by the consideration that the loop phase shift be zero.

The bias for a resonant-circuit oscillator is obtained from an R_gC_g parallel combination in series with the grid, as in Fig. 3. The grid and cathode of the tube act as a rectifier, and if the R_gC_g time constant is large compared with one cycle, the grid-leak capacitor will charge up essentially to the peak grid swing. This voltage across C_g acts as the bias, and the grid is there-

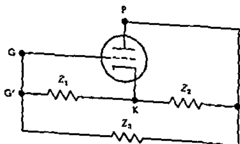


Fig 4 The basic configuration for many resonant-circuit oscillators.

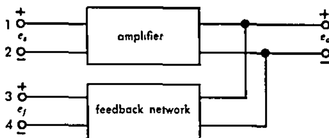


Fig. 1. An amplifier and feedback network not yet connected to form a closed loop.

duced in a signal while being transmitted through a reactive network is invariably a function of the frequency, and there is usually only one frequency at which e_f and e_s are in phase. Therefore, a sinusoidal oscillator operates at the frequency for which the total phase shift of the amplifier and feedback network is precisely zero (or an integral multiple of 2π). The frequency of a sinusoidal oscillator, provided the circuit oscillates at all, is therefore determined by the condition that the loop phase shift is zero.

Another condition, which must clearly be met if the oscillator is to function, is that the magnitude of e_s and e_f must be identical. If the amplifier has a voltage amplification, or gain, A then e_o equals Ae_s. The fraction of the voltage e_o applied to the feedback network is called the feedback factor β. Therefore,

$$e_f = \beta e_o \quad \text{and} \quad e_f = \beta A e_s$$

If e_f is to equal e_s, then βA must equal 1. βA is called the loop gain.

An oscillator will not function if, at the oscillator frequency, the magnitude of the product of the gain of the amplifier and the feedback factor of the feedback network is less than unity. The condition of unity loop gain ($\beta A = 1$) is called the *Barkhausen criterion.*

Referring again to Fig. 1, if βA at the oscillator frequency is precisely unity and the feedback voltage is connected to the input terminals, the circuit will operate with the external generator removed. If βA is less than unity, the removal of the external generator will immediately result in a cessation of oscillations. If βA is greater than unity, 1 volt appearing initially at the input terminals will, after a trip around the loop appear at the input as a voltage larger than 1 volt. After another trip around the loop this larger voltage will be still larger, and so on. It seems, then, that if βA is larger than unity, the amplitude of the oscillations will continue to increase without limit. Of course, such increases in the amplitude can continue only as long as it is not limited by nonlinearity of operation in the active devices associated with the amplifier. Such a nonlinearity becomes more marked as the amplitude of oscillations increases. This onset of nonlinearity to limit the amplitude of oscillations is an essential feature of the operation of all practical oscillators, because all oscillators operate with βA greater than one. The condition that βA equal 1 imposes a single and precise condition of operation, which is not practical in electronic design. Even if the circuit were initially designed to satisfy this condition, it could not be maintained because circuit components (especially vacuum tubes and transistors) change characteristics (drift) with age, temperature, and voltage. Therefore, if the oscillator is left to itself, in a short time βA will become either less than or larger than unity. An oscillator in which the loop gain is exactly unity is an abstraction that is completely unrealizable in practice. A practical oscillator always has a βA somewhat larger than unity (say 5%) to ensure that, with incidental variations in transistor, tube, and circuit parameters, βA does not fall below unity.

Phase-shift oscillator. The phase shift oscillator, Fig. 2, exemplifies the principles set forth above. An amplifier of conventional design is followed by three cascaded arrangements of a capacitor C and a resistor R, the output of the last RC combination being returned to the grid. The phase of the signal is shifted 180° by the amplifier, and the network of resistors and capacitors shifts the phase by an additional amount. At some frequency the phase shift introduced by the RC network is precisely 180°, and the total phase shift around the circuit is exactly zero. At this particular frequency the circuit will oscillate, provided that the magnitude of the amplification is sufficiently large.

The phase shift for the RC network is 180° when

$$f = 1/(2\pi RC\sqrt{6})$$

At this frequency of oscillation β equals $-(1/29)$. For βA to be greater than unity, A must be at least 29. The oscillator then cannot be made to work with a tube like the 12AU7 ($\mu = 20$). It will work with a 12AX7 ($\mu = 100$). The tube employed is often a pentode like a 6AC7 or 6AU6.

The phase-shift oscillator is particularly suited to frequencies from several cycles (per second) to several hundred thousand cycles and so includes the range of audio frequencies. At frequencies in the range of megacycles it has no marked advantage over circuits employing a tuned LC network

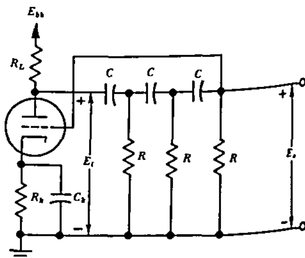


Fig. 2. An RC phase-

Fig 7. Wien-bridge oscillator

(for example, the lighthouse tube) having close spacing between elements are often used

Bridge oscillators. In a bridge circuit the output is in phase with the input at the balance frequency ω_0. Hence, this *circuit may* be used as the feedback network for an oscillator, provided that the phase shift through the amplifier is zero. This condition requires a *two-stage* amplifier. Figure 7 is the schematic diagram of a simple Wien-bridge oscillator. The *frequency* of oscillation is precisely the null *frequency* of the balanced bridge, namely, $f_0 = 1/(2\pi RC)$

The *output* of a balanced bridge is zero when ω equals ω_0, therefore β *and* βA *are* both 0 at this frequency. To *satisfy* the Barkhausen condition ($\beta A = 1$), the bridge must be unbalanced, but in such a way that the phase shift remains zero. *This is* accomplished by making the ratio $R_2/(R_1 + R_2)$ smaller than ⅓. In Fig 7, the coupling capacitors are made large enough so that they introduce no appreciable phase shifts even at the lowest frequencies of operation. The *resistor* R_2 serves both as an element of the bridge and also as a cathode resistor for the first tube. The lower of the two resistors R serves also a dual purpose of bridge element and grid resistor.

Continuous variation of frequency is accomplished by varying simultaneously the two capacitors C (ganged variable air capacitors). Changes in *frequency* range are accomplished by simultaneously switching in different values for the two identical *resistors* R

Practical frequency limits are determined from the *circuit* components. Ganged variable resistors that track with the same precision as do ganged variable air capacitors are not readily available. If variable air capacitors are to be used, they are *necessarily relat*...

frequenc...

used La...

R_3 is als... ...istor cause difficulty because of the possibility that the vacuum tube will block and because a large impedance from grid to ground makes it difficult to shield the grid against stray 60-cycle voltages from the power supply

With the exercise of care, resistors R of the order of 10 megohms may be employed and the frequency pushed as low as two cycles. Also at low frequencies the problem of selecting adequately large coupling condensers becomes more difficult. At the higher frequencies smaller values of resistances R are required. These decrease the impedance looking into the input terminals of the Wien bridge and so increase the loading on the amplifier. Even if the loading is not adequate to stop the oscillation, it will adversely affect the stability of amplitude of oscillation with change of frequency range.

If in Fig 7 the resistor R_2 is replaced by a tungsten-filament bulb, the amplitude is stabilized against variations due to range switching and also those due to the aging of tubes and circuit components. The regulation mechanism introduced by the tungsten bulb automatically changes β to keep βA more nearly constant whenever the value of A should change, as when amplifier loading changes. The resistance of a tungsten *filament increases* with temperature, and the *temperature is in turn* determined by the root-mean-square value of the current which *passes through it*

Other types of bridge networks, such as the twin-T and bridge-T, may be used as feedback elements to form an oscillator. The general principles enunciated above are applicable to these bridge-type oscillators, although the practical details are different.

Crystal oscillators. If a piezoelectric crystal, usually quartz, has electrodes plated on *opposite faces* and if a potential is applied between these electrodes, *forces* will be exerted on the bound charges *within the crystal*. When properly mounted, *deformations* take place within the crystal, and an electromechanical system is formed which will vibrate when properly excited. The resonant frequency and the Q depend upon the crystal dimen-sions...

...ncies ranging from

Fig 8 Electrical equivalent circuit of a piezoelectric crystal

fore driven slightly positive only for a short interval at the peak of the swing. Since the grid base of the tube is traversed in a small fraction of one cycle, the operation is Class C.

When the circuit is first energized, the grid bias is zero, and the tube operates with a large value of transconductance g_m. The loop gain is therefore greater than unity, and the amplitude of oscillation starts to grow. As it does, grid current is drawn, and the bias automatically adjusts itself so that its magnitude equals the peak value of the grid voltage. As the bias becomes more negative, the value of g_m decreases, and finally the amplitude stabilizes itself at that value for which the loop gain for the fundamental frequency is reduced to unity.

The tuned-plate oscillator is only one of many resonant-circuit oscillators, almost all of which have the general configuration illustrated in Fig 4. If it is assumed that the impedances Z are pure reactances X (either inductive or capacitive), then, from the Barkhausen criterion, the circuit will oscillate at the resonant frequency of the series combination of X_1, X_2 and X_3. Also, the loop gain is given by

$$A\beta = \frac{+\mu X_1}{X_2}$$

Since βA must be positive and at least unity in magnitude, then X_1 and X_2 must have the same sign. In other words, they must be the same kind of reactance, either both inductive or both capacitive. Then $X_3 = -(X_1 + X_2)$ must be inductive if X_1 and X_2 are capacitive, or vice versa.

If X_1 and X_2 are capacitors and X_3 is an inductor, the circuit is called a Colpitts oscillator. If X_1 and X_2 are inductors and X_3 is a capacitor, the circuit is called a Hartley oscillator. In this latter case, there may be mutual coupling between X_1 and X_2. If X_1 and X_2 are tuned circuits and X_3 represents the grid-to-plate interelectrode capacitance, the circuit is called a tuned plate, tuned-grid oscillator. Both grid and plate circuits must be tuned to

[illegible]

the inductor L, whose reactance is high compared with X_2. The capacitor C has a low reactance at the frequency of oscillation. However, at zero frequency it acts as an open circuit. Without this

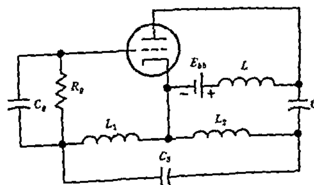


Fig. 5 Hartley oscillator.

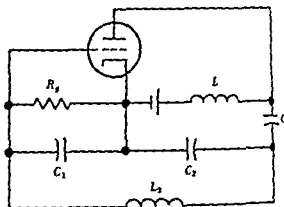


Fig 6 Colpitts oscillator.

capacitor the B supply voltage would be short-circuited by L in series with L_2. The parallel combination of C_g and R_g acts to supply the bias. The circuit operates in Class C, and the grid current charges up C_g to provide the grid bias voltage.

For a low-power oscillator it is possible to use series feed instead of the shunt feed indicated in Fig 5. The B supply is placed between the cathode (ground) and L_2. This eliminates the use of L and C, but L_2 must be insulated from ground for a voltage equal to the B supply plus the peak ac voltage developed across L_2.

A modified form of the Hartley circuit employs mutual coupling between L_1 and L_2 and places C_3 in parallel with L_2.

The practical form of the Colpitts circuit is shown in Fig. 6. This circuit operates in Class C. Capacitor C_1 serves the double purpose of a frequency-determining element and a grid leak.

Electron-coupled oscillator. In this circuit a single pentode tube is used to provide isolation between the generator elements and the output circuit. An example of such an oscillator is a pentode with the Hartley arrangement of Fig 5 connected between cathode, grid and screen and with an output resonant circuit in series with the plate. Since the plate voltage of a pentode has little effect on the plate current, the load is isolated from the oscillator section.

Very-high-frequency (vhf) oscillators. These operate in the range of from a few to several hundred megacycles. The basic configuration of these oscillators is similar to that indicated in Fig 4. However, usually the impedances in the circuit are not lumped elements but are rather distributed (a parallel-wire transmission line or a coaxial cable). These elements are adjusted so that they appear as pure reactances. Sometimes a tuning element, called a butterfly, is used with a vhf oscillator. This element is similar to a variable air capacitor except that the stator plates have holes cut in them in the shape of the wings of a butterfly. As the rotor is turned the inductance (the magnetic energy storage) as well as the capacitance (the electrostatic energy storage) is varied. Hence, tuning over a wide frequency range is possible.

Since the transit time of a [illegible]ectron between the electrodes may be an [illegible] raction of a cycle at these very h[illegible]

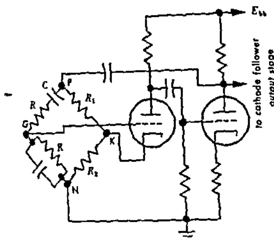


Fig 7. Wien bridge oscillator.

(for example, the lighthouse tube) having close spacing between elements are often used

Bridge oscillators. In a bridge circuit the output is in phase with the input at the balance frequency ω_0. Hence, this circuit may be used as the feedback network for an oscillator, provided that the phase shift through the amplifier is zero. This condition requires a two-stage amplifier. Figure 7 is the schematic diagram of a simple Wien-bridge oscillator. The frequency of oscillation is precisely the null frequency of the balanced bridge, namely, $f_0 = 1/(2\pi RC)$.

The output of a balanced bridge is zero when ω equals ω_0; therefore β and βA are both 0 at this frequency. To satisfy the Barkhausen condition ($\beta A = 1$), the bridge must be unbalanced, but in such a way that the phase shift remains zero. This is accomplished by making the ratio $R_2/(R_1 + R_2)$ smaller than ⅓. In Fig. 7, the coupling capacitors are made large enough so that they introduce no appreciable phase shifts even at the lowest frequencies of operation. The resistor R_2 serves both as an element of the bridge and also as a cathode resistor for the first tube. The lower of the two resistors R serves also a dual purpose of bridge element and grid resistor.

Continuous variation of frequency is accomplished by varying simultaneously the two capacitors C (ganged variable air capacitors). Changes in frequency range are accomplished by simultaneously switching in different values for the two identical resistors R.

Practical frequency limits are determined from the circuit components. Ganged variable resistors that track with the same precision as do ganged variable air capacitors are not readily available. If variable air capacitors are to be used, they are necessarily relatively small in capacity. To attain low frequencies, therefore, large resistances R must be used. Large [illegible]

[illegible] because a large impedance from grid to [illegible]

With the exercise of care, resistors R of the order of 10 megohms may be employed and the frequency pushed as low as two cycles. Also at low frequencies the problem of selecting adequately large coupling condensers becomes more difficult. At the higher frequencies smaller values of resistances R are required. These decrease the impedance looking into the input terminals of the Wien bridge and so increase the loading on the amplifier. Even if the loading is not adequate to stop the oscillation, it will adversely affect the stability of amplitude of oscillation with change of frequency range.

If in Fig. 7 the resistor R_2 is replaced by a tungsten-filament bulb, the amplitude is stabilized against variations due to range switching and also those due to the aging of tubes and circuit components. The regulation mechanism introduced by the tungsten bulb automatically changes β to keep βA more nearly constant whenever the value of A should change, as when amplifier loading changes. The resistance of a tungsten filament increases with temperature, and the temperature is in turn determined by the root-mean-square value of the current which passes through it.

Other types of bridge networks, such as the twin-T and bridge-T, may be used as feedback elements to form an oscillator. The general principles enunciated above are applicable to these bridge-type oscillators, although the practical details are different.

Crystal oscillators. If a piezoelectric crystal, usually quartz, has electrodes plated on opposite faces and if a potential is applied between these electrodes, forces will be exerted on the bound charges within the crystal. When properly mounted, deformations take place within the crystal, and an electromechanical system is formed which will vibrate when properly excited. The resonant frequency and the Q depend upon the crystal dimensions, how the surfaces are oriented with respect to its axes, and how the device is mounted (see PIEZOELECTRIC CRYSTAL). Frequencies ranging from

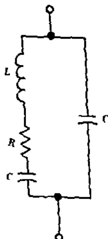


Fig. 8. Electrical equivalent circuit of a piezoelectric crystal

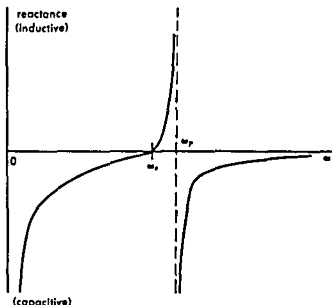


Fig. 9. The reactance function of a crystal (whose resistance has been neglected).

a few kilocycles to a few megacycles and Qs in the range from several thousand to several hundred thousand are commercially available These extraordinarily high values of Q and the extremely stable characteristics of quartz with respect to time and temperature account for the exceptional frequency stability of oscillators using crystals

The electrical equivalent circuit of a crystal is indicated in Fig 8 The inductor L, capacitor C, and resistor R are the analogs of the mass, the compliance (the reciprocal of the spring constant), and the viscous damping factor of the mechanical system Typical values for a 90-kc crystal are an L of 137 henrys, a C of 0.0235 micromicrofarads ($\mu\mu$f), and an R of 15 kilohms, corresponding to a Q of 5500 The dimensions of such a crystal are 30 by 4 by 1.5 mm. Since C' represents the electrostatic capacitance between electrodes with the crystal as a dielectric, its magnitude ($\cong$3.5 $\mu\mu$f) is much larger than C.

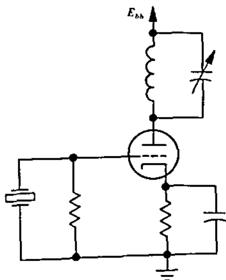


Fig. 10. Crystal version of the tuned-plate, tuned-grid oscillator

If we neglect the resistance R, the impedance of the crystal is a reactance jX whose dependence upon frequency is given by

$$jX = -\frac{j}{\omega C'}\frac{(\omega^2 - \omega_s^2)}{(\omega^2 - \omega_p^2)}$$

where $\omega_s^2 = 1/LC$ is the series resonant frequency (the zero-impedance frequency) and $\omega_p^2 = (1/L)(1/C + 1/C')$ is the parallel resonant frequency (the infinite-impedance frequency). Since C' is much greater than C, then $\omega_p \cong \omega_s$. For the crystal whose parameters are specified above, the parallel frequency is only 0.3 of 1% higher than the series frequency. For $\omega_s < \omega < \omega_p$ the reactance of the crystal is inductive; outside this frequency range it is capacitive, as indicated in Fig 9

Various crystal oscillator circuits are possible. If, in the basic configuration of Fig 4, a crystal is used for Z_1, a tuned LC combination for Z_2, and the capacitance C_{pg} between plate and grid for Z_3, the circuit of Fig 10 results. The crystal reactance, as well as that of the LC network, must be inductive. Hence, the circuit oscillates at a frequency which lies between ω_s and ω_p Since $\omega_p \cong \omega_s$, the oscillator frequency is essentially determined by the crystal and not by the rest of the circuit Figure 10 is the crystal version of the tuned-plate tuned-grid oscillator.

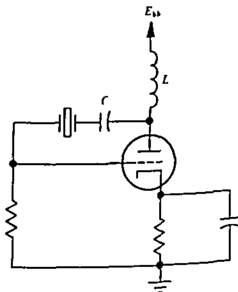


Fig 11 Pierce crystal oscillator

If Z_1 in Fig 4 is grid input capacitance, Z_2 is the plate output capacitance, and Z_3 is a crystal, the result is the circuit of Fig 11, called the Pierce crystal oscillator. This is the crystal version of the Colpitts oscillator The crystal reactance must be inductive. The rf choke L and the blocking capacitor C serve the same functions as they did in Fig 6 This circuit has the merit of not requiring any tuning as one crystal is replaced by another to change the frequency.

The Meacham crystal oscillator is a bridge-type circuit and is indicated in Fig 12 The crystal operates at its series-resonan ency The behavior of this circuit is s of the Wien-

bridge oscillator This circuit has excellent amplitude stability because of the lamp and exceptional frequency stability because of the crystal.

Negative-resistance oscillators. This designation applies to a parallel inductance-capacitance combination placed across a two terminal negative-resistance element If, because of the internal physics of the device, the input current decreases as the input voltage increases the component is said to possess a negative resistance One such device is a tetrode with external terminals considered as the plate and cathode. An oscillator employing a tetrode in this manner is called a dynatron There are a number of semiconductor devices which over a portion of their volt-ampere characteristics possess a negative resistance These elements can be used in oscillator circuits.

A transient in a circuit containing a positive resistance must die down with time because of the losses Hence, an interesting interpretation of the fact that the amplitude first builds up in an oscillator is that during this process the circuit exhibits a negative resistance In this sense, all oscillators might be called negative-resistance oscillators

Heterodyne oscillator. In the heterodyne or beat-frequency, oscillator circuit the voltage from one radio-frequency oscillator is mixed with the output from a similar device tuned to a slightly different

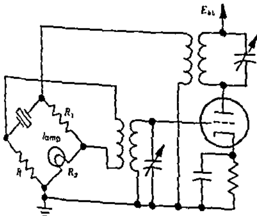


Fig 12. Meacham crystal bridge oscillator

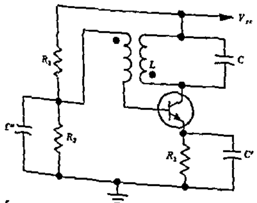


Fig 13 Transistor resonant-circuit oscillator

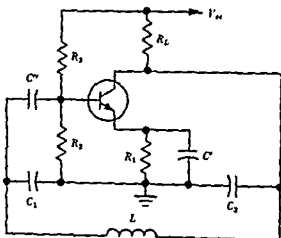


Fig. 14 Transistor Colpitts oscillator

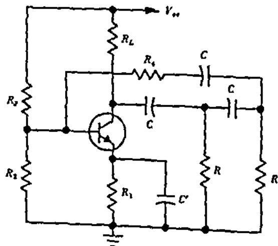


Fig 15 Transistor phase-shift oscillator.

frequency The difference frequency, or beat note, may be varied over the audio or video range by means of a tuning capacitor.

Transistor oscillators. The general theory developed for vacuum-tube *oscillators* is equally valid for transistor sinusoidal generators. In particular, the Barkhausen criteria, the ideas involved in the building up of oscillations, and the limiting of the amplitude due to nonlinearities may be applied to transistor oscillators. Even the specific circuit arrangements using vacuum tubes have their transistor counterparts For example, the transistor resonant circuit oscillator of *Fig 13* is analogous to the vacuum-tube resonant-circuit oscillator of Fig 3. The transistor version of the Colpitts oscillator is given in Fig 14 and the phase-shift oscillator in Fig 15 *See* TRANSISTOR CONNECTION

Microwave oscillators. Special tubes are used for generating waveforms *whose* frequency range lies between a few hundred and several tens of thousands of megacycles *See* MICROWAVE TUBE [J.MI.]

Bibliography: J Millman, *Vacuum tubes and Semiconductor Electronics*, 1958; F. E. Terman and J. M Pettit, *Electronic Measurements*, 2d ed., 1952

Oscillograph

A measurement device for determining waveform by recording the instantaneous values of a quantity such as voltage, as a function of time. It consists of three major components: (1) a primary detector for sensing the instantaneous values of the quantity, (2) the timing system for introducing a time scale on the record, and (3) some means for recording the waveform. *See* WAVEFORM DETERMINATION.

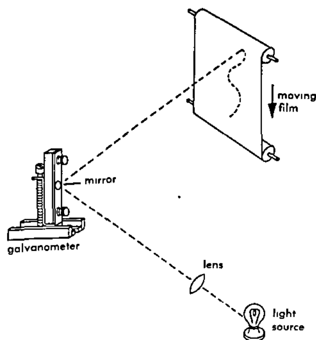
There are two basic forms of oscillographs in common use: the electromagnetic oscillograph and the cathode-ray oscillograph.

Electromagnetic oscillograph. These are either direct-writing oscillographs or light-beam oscillographs. The best known of the direct-writing forms is the electrocardiograph. In this device the primary detector is a galvanometer with a multiturn moving coil. Attached to the moving coil is a pen arm, which traces an ink record on a continuously moving paper chart. The heart beats at about 1 cycle per second (cps) and at this low frequency the pen follows the vibrations nicely and draws a graph in ink that can be read immediately. *See* ELECTROCARDIOGRAPHY; GALVANOMETER.

When the frequency of the wave being recorded by a direct-writing oscillograph exceeds 100 cps, the required speed of the pen is too high for accurate recording.

The light-beam oscillograph, usable up to 500 cps, is a more accurate and more commonly used type of oscillograph. The essential components of a light-beam type magnetic oscillograph are shown in the illustration.

The galvanometer has a moving coil, usually with a single U-shaped turn (bifilar type) but multiturn coils are sometimes used. The bifilar construction contributes to a low moment of inertia and a corresponding high resonant frequency of the moving system (from 3000–10,000 cps in commercial oscillographs). This is desirable, because the resonant frequency must be considerably higher than the highest frequency of the waveform being determined; otherwise the oscillograph would not be sensitive enough to record the waveform.



Bifilar electromagnetic oscillograph with light-beam, photographic-film recording means (*From I. F. Kinnard, Applied Electrical Measurements, Wiley, 1956*)

An optical system provides an inherently high-speed recording means. A tiny mirror attached to the moving coil is the only mass added to the moving system for recording purposes. A beam of light, collimated to a point of light by the lens, is reflected from the mirror onto a photographic film. The lateral position of the point where the light impinges on the film is a function of the position of the mirror and, therefore, the instantaneous value of the current flowing in the moving coil. The photographic film is moved at a known constant speed, the light beam therefore traces the waveform on the film. Development of the film is required before the record is visible.

Cathode-ray oscillograph. The low-frequency limitations (500 cps) of the relatively large masses of the electromagnetic oscillograph element are eliminated in the cathode-ray oscillograph. This device is a cathode-ray oscilloscope combined with a camera that makes a photographic record of the screen image. (*See* OSCILLOSCOPE, CATHODE-RAY.)

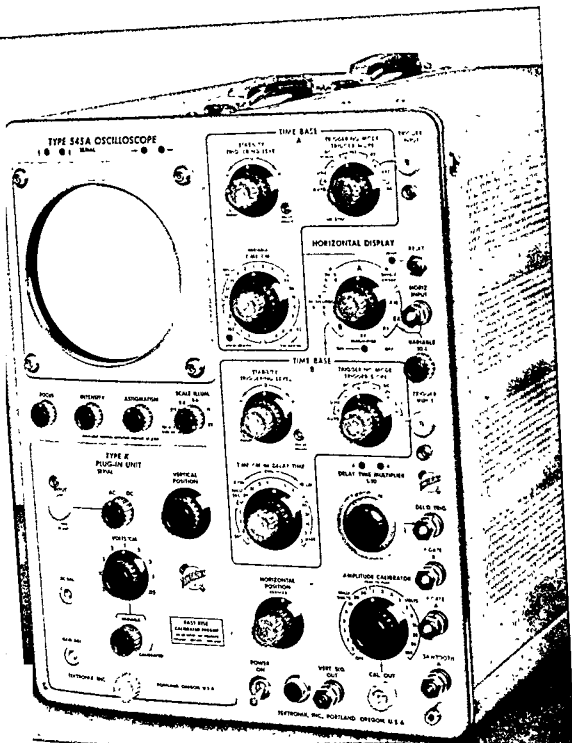
The terms cathode-ray oscillograph and cathode-ray oscilloscope are commonly interchanged. Strictly speaking the oscillograph contains means for producing records, whereas the oscilloscope does not. [I F K]

Bibliography: H. Buckingham and E. M. Price, *Principles of Electrical Measurements*, 1955; I. F. Kinnard, *Applied Electrical Measurements*, 1956

Oscilloscope, cathode-ray

An electronic instrument which produces a luminous plot on a fluorescent screen showing the relationship of two or more variables. In most cases it is an orthogonal (x,y) plot with the horizontal axis being a linear function of time. The vertical axis is normally a linear function of voltage at the signal input terminal of the instrument. Because transducers of many types are available to convert almost any physical phenomena into a corresponding electric voltage, the oscilloscope is a versatile tool in all forms of physical investigation.

The primary advantage of a cathode-ray oscilloscope over other forms of plotting devices is its speed of response. Commercially available instruments in the general-purpose category can display frequencies as high as 100 megacycles (Mc) while special high-speed oscilloscopes can respond as high as 2000 Mc. The horizontal linear time axis of one general-purpose oscilloscope may be varied in 25 calibrated ranges from a slow speed of 10 cm in 50 sec to a high of 10 c[illegible] microseconds (μsec). The same oscill[illegible] on photo-



ig 1 Typical commercial wide-band (dc–30 Mc) oscilloscope with precision sweep delay and plug-in preamplifiers (Tektronix, Inc)

graphic film a single trace at a rate of 250 cm/μsec

In its normal form the cathode-ray oscilloscope is made up of five basic elements (Fig 1)

1 The cathode-ray tube and associated controls in focus, intensity or brightness, and astigmatism.

2 The vertical or signal amplifier (plug-in unit in Fig 1) and its associated devices such as input terminal, attenuators, position control, and ac or dc amplifier operation selector

3 Horizontal-axis time-base circuits, frequently called the sweep generator. Included in this group are the sweep base control, trigger or synchronizing circuits, and usually a circuit for turning on the cathode-ray tube beam only when the sweep is going in the left-to-right direction on the screen

4. Auxiliary facilities, such as amplitude or time calibrators and repetition rate generators.

5. Power supplies furnishing the correct operating voltages for the above circuits.

Cathode-ray tube. The central component in a cathode-ray oscilloscope is the cathode-ray tube itself. This is frequently referred to as the CRT. In its simplest form it consists of an evacuated glass container with a fluorescent screen at one end and a focused electron gun and deflection system at the other end.

Either magnetic or electric fields may be used for focusing and deflection. Oscilloscopes almost always use electric fields for both functions. Electrostatic focusing is used largely because of its convenience and simplicity. Electrostatic deflection is almost universally used for oscilloscopes because it is capable of superior high-frequency response. The only magnetically deflected oscilloscopes are those using TV picture tubes for application where 100-kc bandpass or less is adequate, but a large image is needed.

The CRT designer is faced with four primary technical objectives: (1) high deflection sensitivity; (2) a bright image for ease of observation and photography; (3) small spot size relative to the image area; and (4) accurate deflection geometry All of these are related in various ways so that each tube is the result of compromises which seem least undesirable to its designer *See* CATHODE-RAY TUBE

Signal amplifier. The signal is usually applied to the vertical axis of the oscilloscope; thus this amplifier is frequently called the vertical amplifier. Signals commonly observed by means of the cathode-ray oscilloscope vary in amplitude from microvolts to kilovolts and in frequency from dc to many megacycles. The deflection sensitivity of common types of cathode-ray tubes lies in the range of 0 01–0 2 cm/volt, with from 4–10 cm of deflection available. Thus, for many purposes, it is necessary to amplify the signal in order to get sufficient deflection on the tube for accurate observation or measurement. The prime requirement of this amplifier is that it must produce an amplified replica of the signal applied to its input with a minimum of distortion or variation in wave shape This requires an amplifier with adequate frequency and phase response, in addition to a very linear transfer characteristic

Most modern oscilloscopes use dc amplifiers, that is, amplifiers whose low-frequency response extends to zero. This feature is valuable for several reasons First and most obvious, slowly changing phenomenon can be accurately observed Second, the relation of a waveform to essential dc reference levels can be observed. *See* AMPLIFIER; DIRECT-COUPLED AMPLIFIER; VOLTAGE AMPLIFIER

Signal delay networks. In order to view widely spaced pulses, especially those whose spacing is not uniform, it is necessary to trigger the time sweep directly from the pulse being observed. Since the sweep takes a small but finite time to get started, it is necessary to delay the signal for a slightly longer time so it will appear away from the extreme edge of the screen. This is accomplished by inserting a delay network (adjusted by control knobs as shown in Fig. 1) in the signal channel after the trigger take-off point.

If this delay network is to transmit the signal without waveform distortion, it must be carefully designed and constructed. Three types are used

1. Coaxial cable, which has little signal distortion but is bulky—200 ft would be necessary for 0 25 μsec delay.

2. Continuously wound delay lines, usually on a ferrite core. These are compact and can be made in convenient impedances but have rather high attenuation at frequencies of 15 Mc or more Coupling networks on both ends are usually needed to eliminate reflectors from the capacitances of the input and output circuits

3. Distributed constant networks This type is used in commercial oscilloscopes with bandwidths as high as 50 Mc. They are usually of the M-derived type with an adjustable capacitance in each section. With push-pull amplifiers, a balanced network is used, having two sets of inductances with the capacitors connected between points of equal delay. This type of network may consist of as many as 90 sections. When properly adjusted, the aberrations on a square pulse are less than 1% of the pulse amplitude.

Attenuators and gain controls. To obtain a suitable image size on the cathode-ray tube, it is necessary to have a convenient means of varying the amplifier gain or sensitivity. Two methods are used in modern oscilloscopes. First, a compensated step attenuator is placed in front of the amplifier. There are usually two or three steps per decade with a total attenuation such that signals of several hundred volts can be observed. So that the attenuator will have the same attenuation for all frequencies it usually consists of resistance and capacitance dividers in parallel. When both dividers have the same ratio, the attenuation is independent of frequency If only the resistive section were used the stray shunt capacitance would increase the attenuation with frequency. These strays become a part of the capacitance divider and are rendered harmless.

The second gain control is usually an uncalibrated, variable one having a range of 3 to 1 or less This fills in between the steps of the fixed attenuators It is usually a low-impedance control so that high-frequency compensation is not necessary

Differential input amplifiers. Most oscilloscope signal amplifiers have one terminal, usually called the ground terminal, connected to the chassis of the instrument This is normally satisfactory because most waveforms being observed have a common or ground reference also Many times, however, it is necessary to observe the waveforms between two points, neither of which is grounded For this purpose balanced, or differential, amplifiers are needed.

The output of this type of amplifier is proportional to the algebraic diff[illegible]e of the signals applied to its two input [illegible] signals com[illegible] m[illegible] both termin[illegible] [illegible]ut [illegible]

ferences between the terminals are amplified In gh-sensitivity differential amplifiers the amplificaon of the desired signal may be as much as 10 000 nes the common mode signal. This property often akes it possible to observe a small signal in the resence of large interfering signals. A differential nplifier is not shown but can be plugged in in lace of the vertical amplifier shown.

Distortion in an oscilloscope. For purposes of his discussion, distortion is said to occur if the vaveform on the screen is not a replica of the input vaveform except, of course, for a change in scale. Among the causes for distortion are the following:

1 The cathode-ray tube may have different deflection factors in various portions of the screen or may not have orthogonal axes

2 The amplifier may not have a linear relationship between input and output amplitudes

3 The frequency response of the oscilloscope may be inadequate at either or both ends of the frequency spectrum of the signal being observed.

4 All components of the signal may not arrive at the deflection plates at the same time.

The remedy for the first cause is obviously a cortly designed and built cathode-ray tube. In the ond case, the designer should first use tubes with most linear transfer characteristics, operated optimum grid bias, screen, and plate voltages rther reduction in amplitude distortion can then obtained by using balanced or push pull cirits and negative feedback when possible

If the high frequency response is inadequate, he slope of the steeply rising and falling portions f the waveform will be decreased When an ininitely steep wave front, or step, is applied to a ircuit having a finite bandpass, it will rise or fall according to the following approximate formula:

$$T_r = \frac{0.35}{f}$$

T_r is rise time in microseconds from 10 to 90% amplitude, f is frequency in megacycles on the bandpass curve where response falls to 70% of the midfrequency amplitude

Thus, a zero-rise time step displayed on a 1-Mc bandpass oscilloscope will appear to have a rise time of 0.35 μsec. The bandpass necessary to observe a signal with the desired accuracy is obtained from a formula relating the resultant rise time T_r for a signal of finite rise time T_{r1} going through an oscilloscope of known rise time T_{r2} This is simply

$$T_r = \sqrt{T_{r1}^2 + T_{r2}^2}$$

Thus, a pulse with a 1-μsec rise in passing through a 1-Mc oscilloscope would appear to have a rise time of [illegible]

[illegible] cycle, 10% response will cause a 10% slope in a 0.001 sec flat portion of the waveform.

The distortion caused by various frequency components arriving at different times is frequently called phase distortion. This causes overshoots or spurious damped oscillation following sudden steep portions of a waveform. The analytical problem of determining the effect of phase distortion on a given waveform is quite complicated, but fortunately the oscilloscope designer has a simple and direct way to observe it in practice This is done by observing the response to a clean, sharp amplitude step whose rise time is short compared with that of the oscilloscope under test. The necessary adjustments to the high frequency compensating circuits are then made until the waveform on the cathode-ray tube has the desired precision

Horizontal sweep and synchronization. The most useful oscilloscope presentation is that having a linear horizontal time axis accurately synchronized with the signal being observed. Prior to 1946, most oscilloscopes except for special purpose pulse monitors and synchroscopes used a sweep generator of the recurrent type, which is one that continues to operate in the absence of synchronizing signals The circuit is adjusted so that its natural frequency is slightly lower than an integral fraction of the signal frequency. When the signal frequency is then injected into the circuit, the sweep is caused to terminate at a fixed point on the signal waveform, immediately returns to its initial value, and starts a new sweep For closely spaced signals, such as sine waves, square waves, and the like, this method is simple and satisfactory. For observation of pulses or other widely spaced waveforms, it cannot be used Consider the case of a 1-μsec pulse at a repetition rate of 1000 pulses/sec. The sweep would operate at the pulse repetition rate and be almost 1000 μsec long A 1-μsec pulse would be hardly visible on such a sweep and certainly no detail could be observed.

Triggered sweep is the solution to this problem In this circuit the sweep is inoperative except when started by a trigger signal When the sweep is completed the circuit returns to its original state and awaits another trigger With this circuit there is no necessary relationship between the repetition rate and sweep speeds so that sweep-speed controls may be varied at will without affecting the synchronization.

In modern oscilloscopes, the sweep generator usually consists of the following elements: (1) trigger selector and amplifier; (2) sweep waveform generator; (3) sweep amplifier; and (4) CRT unblanking circuits

Trigger selector and amplifier If the oscilloscope is to provide a stable image, each sweep must start at the same point on the signal waveform The information needed by the sweep generator to accomplish this may come from several sources: first and most useful, from the signal waveform itself; second, from a separate waveform which has an accurate time relationship with the signal, for example the synchronizing signals in a television system; and third, from the power-line frequency, because many waveforms such as those found in power supplies are accurately related to it A selector switch is frequently provided to enable conven-

ient choice of these sources (trigger-mode adjusting knobs in Fig. 1). These trigger signals may be too small to activate the sweep generator; thus an amplifier is usually provided. This amplifier is frequently of the regenerative type whose output is always a rectangular wave of fixed amplitude regardless of the shape of the input signal. A single steep portion of this wave provides a sharp definite starting signal for the sweep generator. The sweep amplifier frequently provides facilities for causing the sweep generator to start at any selected portion of the waveform.

Sweep waveform generator To provide a linear time axis, it is necessary to apply a sawtooth voltage waveform to the CRT horizontal plates. This waveform is one which starts at a fixed voltage, rises or falls at a linear rate to a second fixed voltage, and then returns to the first reference to repeat the cycle The sweep waveform is generated by the charge and discharge of a capacitor The linear portion used on the sweep is obtained by charging or discharging the capacitor with a constant current. When the linear portion is completed, the capacitor is brought back to its original voltage as rapidly as possible. In order to make sure that the linear portion starts from a stable reference on each sweep, auxiliary circuits are frequently used which prohibit the trigger signals from reacting the sweep generator until the timing capacitor has had adequate time to stabilize at its initial reference voltage. This circuit is usually called the hold off *See* SWEEP GENERATOR

Sweep amplifiers Two functions are usually provided by the sweep amplifier. First, it amplifies the sweep generator waveform to that required to deflect the CRT beam, and second, it provides balanced signals to the deflection plates of the CRT Balanced signals imply that as one plate goes in a negative direction, the other goes positive an equal voltage. This is essential for accurate image geometry.

In order to accomplish its function, a sweep amplifier must have a very linear transfer characteristic, high gain stability, and a frequency response adequate to amplify the linear portion of the sweep waveform uniformity from the slowest to the fastest sweep rates

CRT unblanking circuits. The CRT beam is normally turned on only during the linear portion of the sweep waveform. This is especially necessary in pulse observation because the space between pulses may be several hundred times the sweep length. If the beam were not turned off between sweeps, it would rest in a bright spot at the edge of the screen, causing much extraneous light, and would probably damage the screen.

Time-interval measurement. There are several methods of measuring time intervals.

Calibrated sweeps. Practically all laboratory oscilloscopes have calibrated horizontal sweeps so that time interval measurements may be read directly from a graticule over the cathode-ray tube screen. Probable error of this method ranges from 1 to 10%, depending on the precision of the oscilloscope and adequacy of maintenance and calibration. This method combines convenience with an ability to cover a wide range of time intervals. One popular oscilloscope has 25 ranges accurate to within 3% covering a range of 5 sec/cm to 0.02 μsec/cm.

Precision sweep delay. For greater accuracy, some form of expanded scale is needed This may be accomplished by an accurately calibrated sweep-delay circuit. To use this method, the oscilloscope sweep is set to a speed so that adequate resolution of the waveform is available. The sweep-delay dial (usually having 1000 divisions) is then turned so that the start of the waveform being measured is under the center mark of the screen. The dial reading is noted and the dial turned so that the end of the waveform is now under the center mark The time interval is thus the difference of the two dial readings. This method is capable of accuracy within 1% or better.

Time marker. Another method which is used principally in connection with radar range measurement and television uses a synchronized precision marker generator. This is usually an accurate sine-wave oscillator started and stopped by the sweep generator. The sine waves are put through shaping circuits which produce a sharp pulse or pips at the same point on each cycle. These pips are usually applied to the cathode-ray tube so as to brighten the trace once during each cycle of the oscillator These bright pips are referred to as time markers. The advantage of this method is that its accuracy is dependent on the calibration of a sine-wave generator rather than the usually less stable sweep generator and cathode-ray tube circuits. If the time mark pips are sharp and the waveform being measured has steep rises and falls, the reading accuracy of this system is excellent, because a small horizontal displacement of a pip will move it a large vertical distance on the waveform Time-mark generators are not widely used on general-purpose oscilloscopes for several reasons First, many ranges would be needed, each with a different time between marks This means a large number of components to be switched Second, on the faster sweep ranges the pips must be very short. On a 0.02 μsec/cm sweep, the pip should be 0.005 μsec or less. This requires a bandwidth for the circuits carrying the pip of about 200 Mc.

Dual-trace oscilloscopes. Frequently it is desirable to compare two waveforms on the face of a single cathode-ray tube. For this purpose several methods are used.

Dual-trace amplifier and single-gun CRT. This method uses a circuit frequently called an electronic switch The oscilloscope amplifier is switched electronically between the two signals under observation. If this switching is made to take place in the interval between sweeps when the beam is blanked out, the presentation on the screen is indistinguishable from a dual beam oscilloscope This method of presentati[illegible] impler and less expensive than a true [illegible] loscope and als[illegible] vides more

een the two signals because time-base errors are mmon to both waveforms

Split beam CRT. This oscilloscope has been ipular in England for many years It uses a single ectron beam with a splitter plate in front of the ial aperture The beam is divided into two parts, ch receiving separate vertical deflection Both rts of the beam, however, pass through a single ur of horizontal deflection plates

Dual gun CRT. This method is similar to the lit beam method but has the additional flexibility separate brightness, focus control, and separate ilanced vertical deflection plates, Dual-trace amlifiers can be provided in both vertical systems so at four traces are available, if desired

The most versatile arrangement is that with two parate oscilloscopes presenting their waveforms a common screen for convenient comparison A -30 Mc instrument with versatile sweep facilities available One sweep is available to provide a recision delay for the other sweep Thus, an entire aveform and a highly magnified portion of it can e shown simultaneously on the same screen.

High-speed oscilloscopes. As technology progresses, the need for a higher degree of time resolution becomes increasingly important When signal components reach into the hundreds or thousands f megacycles, conventional amplification becomes mpractical. Great care must be taken to avoid mismatches in the signal channel because the resulting reflections would distort the signals being bserved. Very fast sweeps must be provided so hat rise times of millimicroseconds or less may be measured The problem of taking a trigger from the signal being observed without distorting or seriously attenuating it is very difficult To make matters even more difficult, it is frequently necessary to have sufficient light output from the CRT so that single ...

spot trave.

light This

which red

deflection sensitivity increases as the scan angle decreases, most millimicrosecond (mμsec) oscilloscopes have small deflection areas with a correspondingly small spot so that resolution is maintained In these small display areas, the usual measure of sensitivity, volts/cm, becomes meaningless. A more significant measure is to use volts/trace width This is termed sensibility

The problem of transit time of the electron beam through the deflection plates is solved in most cases by some form of a traveling-wave system Here the deflection plates are broken up into a number of segments and arrangements made so that the signal travels from one segment to the next at the same speed as the electrons in the beam Thus any electron is deflected by the same signal component throughout its entire time in the deflection system

Figure 2 shows a cutaway drawing of a commercial traveling-wave CRT having the following performance: (1) vertical (TW) sensibility 0 03 volt/trace width; (2) maximum vertical scan 0 4 (3) vertical frequency response of approximately 2000 Mc; (4) horizontal sensibility 0.2 volt/trace width; (5) horizontal scan 0 6 in.; and (6) writing speed 10^{11} trace widths/sec.

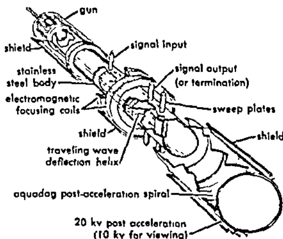


Fig. 2 Cutaway drawing of a commercial traveling-wave deflection cathode-ray tube. (Edgerton, Germeshausen, and Grier, Inc)

A different approach is possible when the signal to be observed is a repetitive one. This involves an amplitude-sampling technique using a short gating pulse, less than 1 mμsec Samples are taken at a slightly later time at each recurrence of the signal. These samples are amplified and lengthened in time and displayed on a much slower sweep. The time resolution of this type of instrument is limited by the length of the gating pulse and the accuracy with which successive pulses may be positioned along the signal. It also has relatively high sensitivity, limited by random noise. The limitation of this type is its inability to observe single transients and the slowness of the presentation for low repetitive rates. For example, a 60-cycle signal would need 2 sec to produce a trace made up of 120 samples The performance specifications of a commercial sampling oscilloscope are as follows: (1) rise time, 0 4 mμsec, (2) vertical sensitivity, 0.4 cm/mv, and (3) maximum apparent sweep rate, 0 04 mμsec/cm.

Electrical measurements. In addition to providing a visual indication of waveform, the modern oscilloscope provides an accurate method of electrical measurement.

Measurement of frequency. If the oscilloscope has a calibrated horizontal axis or time base, frequency may be measured by reading the time necessary for one complete cycle and inverting the result. The accuracy is essentially that of the time-base calibration, whose errors are usually between 1 and 10%. For greater accuracy, the usual method is to substitute a variable calibrated sinusoidal oscillator for the time base and adjust it until a stationary ... discussion of

... A common ... relative phase of two signals of the same frequency is to apply them to the two axes of the oscilloscope and compute the

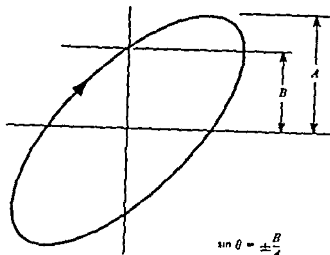


Fig 3. Measuring phase between two signals by means of Lissajous figure.

answer by means of the method shown in Fig 3 The sine of the phase angle θ between the two signals is given directly by measuring the relative heights A and B as indicated in the figure Any suitable scale, such as inches, may be used.

$$\sin\theta = \pm B/A$$

A more convenient method is to use a dual-beam oscilloscope having a common calibrated time base The phase angle can be read directly from the pattern on the screen.

Measurement of voltage Complementing their calibrated sweeps, modern laboratory oscilloscopes have directly calibrated amplitude scales. Thus the voltage difference in dc or peak-to-peak volts is easily read from the calibrated graticule This method is made possible by the development of highly stable amplifiers

To aid in maintaining the accuracy of calibration, many oscilloscopes have built-in precision reference voltage sources These are usually square-wave generators whose limits are set by precision dc voltages and voltage dividers

An oscilloscope always measures in peak-to-peak volts. If the form factor is known for any particular waveform, the peak-to-peak may be converted into rms, average, or any other voltage system

Photography of oscilloscope patterns. Oscilloscope patterns are frequently photographed for two reasons, first, to preserve a repetitive image for future study and comparison, and second to record for study single brief transient waveforms which could otherwise not be studied in detail. The first case is relatively easy because an exposure of sufficient length can be made so that wide-aperture lenses and high-intensity oscilloscopes are not necessary. A typical exposure might be 1 sec at $f/8$.

For photography of all but the faster single transient, cameras using the Polaroid-Land system are popular. A finished positive paper print is available about 1 min after the exposure is made If for some reason a good picture is not obtained, additional exposures can be made until a good one is obtained.

In order to photograph single transients at speeds of 25 cm/μsec or more, conventional 35-mm miniature cameras using wide-aperture lenses ($f/1.4$ or $f/2$) and high-speed films are frequently used

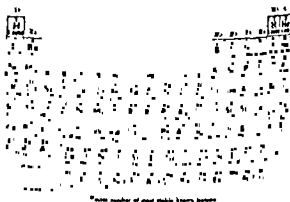
High contrast, rather than a good tonal range, is wanted; therefore extended development in a high-energy developer is usual to obtain maximum film speed. Exposure is less critical than in normal photography for the same reason. [B I]

Bibliography: I. A. Greenwood, J. V. Holdam, and D. MacRae, *Electronic Instruments*, vol 21, 1948; I A. D Lewis and F. H. Wells, *Millimicrosecond Pulse Techniques*, 1954; O. S Puckle, *Time Bases*, 2d ed., 1951; T. Soller, M. A Starr, and G. E. Valley, Jr. (eds.), *Cathode ray Tube Displays*, vol. 22, 1948.

Osmium

Chemical element number 76, osmium, Os is a hard white metal of rare natural occurrence

Uses. Osmium tetroxide is used for the hydroxylation of double bonds in the synthesis of certain organic compounds Particular applications of this reaction occur in the synthesis of cortisone. Osmium tetroxide is also used as a stain for tissue in microscopy. Osmium is an excellent hydrogenation catalyst. It is used in the alloy osmiridium for making pen points



Chemical and physical properties. The atomic number of osmium is 76, atomic weight 190.2, density 22.47 g/cm³, melting point higher than 3000°C, boiling point 5500°C and electrical resistivity 9.5 microhms per centimeter at 0°C. Radioactive isotopes of the following mass numbers are known 182 183, 185, 191, 193, and 194 The stable isotopes of osmium have the mass numbers 184, 186, 187, 188, 189, 190 and 192 The natural abundances of these are respectively 0.018, 1.59 1.64 13.3 16.1, 26.4, and 41.0% of the element. Although osmium is very refractory, even at room temperature, a blue oxide film is formed on the metal surface. When heated in air it oxidizes readily to osmium tetroxide, which is volatile and very poisonous This property complicates the analysis and refining of osmium since losses due to volatilization may be inadvertently encountered is readily soluble

hot nitric acid When fused in alkaline oxidizing ...es, water-soluble osmates, OsO_4^{2-}, are formed ...mium forms compounds in which it has valences 2+, 3+, 4+, 6+, and 8+. The chemistry of ...mium is very complicated because of the many ...lences exhibited by the element and the tendency ...each of these to form numerous complex ions.

Metallurgical extraction. After osmium has been ...rought into solution, it may be distilled from ni...ic acid as the tetroxide. Often solids containing ...smium can be roasted in air to volatilize the te...roxide which is then absorbed in an alcoholic caus...c solution The resulting osmate solution may be ...recipitated as the sulfide or neutralized and pre...ipitated as the hydroxide Either precipitate is ...hen reduced in hydrogen to yield the metal

Principal compounds. Osmium tetrachloride, $OsCl_4$, is a black solid which is insoluble in non-oxidizing acids, it is made by treating osmium with chlorine at 700°C. Osmium tetroxide, OsO_4, is a very pale yellow crystalline solid with a melting point of 40°C and a boiling point of 130°C. It is the most important osmium compound, and is made by oxidizing the metal with air, nitric acid, or sulfuric acid This very poisonous compound is soluble in water and carbon tetrachloride It is a powerful oxidizing agent When a potassium hydroxide solution of the tetroxide is treated with alcohol, the osmium is partly reduced, and slightly soluble violet red crystals of potassium osmate, $K_2OsO_4 \cdot 2H_2O$, are precipitated The dihydrate of osmium dioxide, $OsO_2 \cdot 2H_2O$, is made by neutralizing an alcoholic sodium hydroxide solution of the tetroxide It is a brown or blue black insoluble solid. The solid hexachloroosmic acid, H_2OsCl_6, has not been definitely isolated However, when the tetroxide is refluxed in h... [illegible]

amm... [illegible]

(NH... [illegible]

droge... [illegible]

An... [illegible]

as the tetroxide from nitric acid Hydrated osmium dioxide may then be precipitated reduced in hydrogen, and the resulting osmium metal may be weighed The red color developed by the reaction of thiourea with osmium distillates is used for the colorimetric deter... [illegible]

For ... [illegible]

allurgi... [illegible]

IUM; I... [illegible]

[E A HA]

Osmoregulatory mechanisms

The maintenance of an optimal and constant level of osmotic activity of the fluid within and around the cells considered to be most favorable for the initiation and maintenance of vital reactions in the cell and for maximal survival and efficient functioning of the entire organism

Evolution. Practically all living cells function in a fluid environment. This includes isolated unicellular forms such as *paramecia* and amebas as well as cells comprising tissues in air breathing terrestrial animals Thus the ionic composition and osmotic activity of extracellular fluids have biological significance with respect to survival of the living cells The fluid environment of simple unicellular forms of life consists of the ocean, lakes, or streams in which the forms are found, while that of the complex animal forms consists of the fluid media enclosed by the various compartments of the body

Presumably, life originated in the primitive ocean, the salinity of which was lower in prehistoric than in modern times The first unicellular forms of life remained in a state of osmotic equilibrium with their dilute sea water surroundings; that is, they had no osmoregulatory devices. In the course of evolution, unicellular and multicellular animals migrated from the sea to fresh-water streams, and eventually to dry land. Survival of such forms was associated with maintenance of constant osmotic activity of their body fluids through evolution of osmoregulatory devices while fluids decreased.

Changing osmotic conditions occasioned the evolution of special cells and tissues which permitted retention or exclusion of the proper amount of water and solute for the animal

Biological mechanisms. The actions of osmoregulatory mechanisms are (1) to impose constraints upon the passage of water and solute between the organism and its surroundings and (2) to accelerate passage of water and solute between organism and surroundings The first effect requires a change of architecture of membranes in that they become selectively permeable and achieve their purpose without expenditure of energy. The accelerating effect apart from requiring a change of architecture of cell membranes, requires expenditure of energy and performance of useful osmotic work. Thus, substances may be moved from a region of low to a region of higher chemical activity. Such movement can occur in opposition to the forces of diffusion, of electric field, and of pressure gradient which act across the cell membrane It follows that there must be an energy source derived from the chemical reactions of cellular metabolism and that part of the free energy so generated must be stored in molecules driven across the membrane barrier. Active transport is the modern term for such processes The manner whereby chemical energy can be transferred into osmotic energy has not yet been determined Some examples to illustrate osmoregulation follow

Fresh-water fish and frogs can pump water out of the body via the urine. The survival in dilute fluids is thereby accomplished The salt-water fish pumps salt from its body to the sea across the gills This removes the salts of ingested ocean water, thus permitting survival with brine as the source of water. Of interest is the absence of glomeruli in the kidneys of some marine forms like the toadfish and the goosefish. Such aglomerular kidneys excrete smaller volumes of urine than do glomerular kidneys and possess, in their tubular cells, transport systems for the excretion of salt. The alba-

Summary of osmotic performance in various animals*

Osmotic characteristics	Principal mechanisms	Examples
Osmotic adjustment	No volume regulation	Marine invertebrate eggs, *Phascolosoma*
	Volume regulation	Marine mollusks, *Maja*, *Nereis pelagica*, *N cultrifera*
Limited osmoregulation	Low permeability, salt reabsorption in nephridia?	*N diversicolor*
	Water storage	*Gunda*
Fair osmoregulation in hypotonic media	Selective absorption of salts from medium, kidney reabsorption or secretion, low permeability	*Carcinus*
Regulation in hyper- and in hypotonic media except at extremes	Unknown	*Uca*
Unlimited regulation in hypotonic media	Hypotonic copious urine, salt reabsorption or water secretion, low surface permeability	Crayfish, fresh-water teleosts, Amphibia
	Water impermeability	Fresh-water embryos
Maintenance of hypertonicity in all media	Urea retention	Elasmobranchs
Regulation in hypertonic media	Extrarenal salt excretion, low water intake	Marine teleosts
	Unknown	*Artemia*
Regulation in moist air	Low skin permeability, salt absorption from medium, salt reabsorption in kidney	Earthworm, frog
Regulation in dry air	Impermeable cuticle, hypertonic urine	Insects
	Hypertonic urine, water reabsorption in kidney	Birds and mammals

* From C L Prosser et al (eds), *Comparative Animal Physiology*, Saunders, 1950

tross and penguin, by secreting an extremely hypertonic nasal fluid, can survive with sea water as the sole source of water, a unique biological advantage for a terrestrial animal. *See* SALT GLAND

The table presents a zoological classification of various modes of osmotic defense in several phyla of animals.

Theoretically, the requirements for efficient osmoregulation would be (1) development of active transport mechanisms for solute and for water, (2) development of barriers to free diffusion of solute across membranes or cells, that is, selective permeability characteristics, and (3) the integration of such mechanisms into appropriate organ systems in intact animals. This includes the development of complex controlling systems for osmoregulatory organs. For example, the posterior-pituitary neurohormonal system regulates urinary flow and concentration in birds and mammals. *See* URINARY SYSTEM.

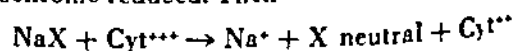
TRANSPORT PROCESSES

A stringent definition of active transport of ions requires that a net amount of material must be moved unidirectionally across a biological barrier that is, either a membrane or cell, against the forces of diffusion, against the forces of an electrical field, and even against the force of a hydrostatic pressure gradient. For individual ions, such a movement is said to be against the electrochemical po[illegible] defined thermodynamically [illegible] An equally [illegible] er transport would require movement of water against gradients of hydrostatic pressure and of chemical potential. In the most general sense, active transport of a substance means its movement against the free energy gradient; therefore, the process requires a source of free energy from cellular metabolism. Exactly how metabolic energy is funneled into a cellular mechanism capable of doing osmotic work is unknown Various hypotheses have been proposed to account for the transport of solutes across cellular membranes. The following is a brief description of the elements of some of the popular schemes for solute transport.

Oxidation-reduction systems. The elements of such a scheme are that (1) there is the presence of oriented electron transporting reaction in the membrane, (2) the reaction results in movement of a given ion, at a single site from a region of low to a region of high electrochemical activity, with generation of an electromotive force; and (3) the electric field so generated transports another ion of opposite charge in the same direction Alternatively the field could move a similarly charged ion, at a separate site in the membrane, in a direction opposite to that of the primary ion movement. In either case, the laws of electroneutrality are satisfied for the whole system, while osmotic work is done on the transported particles The active transport of sodium (Na^+) ion with passive transport of chloride (Cl^-) ion across the frog skin can be described in stepwise manner using the assumptions of an oxidation-reduction mechanism.

Figure 1 illustrates schematically the operation of an oxidation-reduction mechanism designed for the active transport of Na^+ ions.

The reaction in the membrane is oriented, by means of unspecified constraints, at the interface across which the actively transported ion is ejected Thus Na^+ ion enters the cell membrane by combining electrostatically with a carrier substance X^- to yield undissociated NaX The undissociated complex diffuses across the membrane wherein the oriented reaction is between cytochrome-oxidized and cytochrome-reduced. Then



$$NaX + Cyt^{+++} \rightarrow Na^+ + X \text{ neutral} + Cyt^{++}$$

and the free Na^+ ion formed is ejected from the cell as the electron moves in the opposite direction and reduces the o[illegible]hrome To sat-

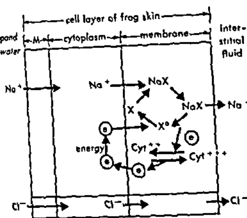


Fig 1. The transport of Na^+ and Cl^- from pond water, 0.01 *M* NaCl, across the frog interstitial fluid, 0.10 *M* NaCl

isfy electroneutrality, a negative ion (Cl^-) must be transported at a site spatially separated from that of cation transport. A separate electron donor substance must be present to convert X neutral to X^- n, a process requiring metabolic energy. Although never proven rigorously, the assumptions of this scheme may be used to account for many osmoregulatory processes in cells across which a measurable electric field exists.

Forced ionic exchange and carrier systems. Some investigators believe that electrical potentials across cell membranes are merely diffusion potentials secondary to the ionic gradients produced by a carrier-type of transport. They postulate the existence in a membrane of a charged substance, say X^-, in the case of cation transport. At one interface, the carrier or the mechanism ejects Na^+ ion from the cell and simultaneously injects K^+ ion into the cells. This is called a forced ionic exchange process and produces no electrical potential. At another site of the membrane, there are pores filled with water. According to the K. Meyer, J. Sievers and T. Teorell concept, there are fixed negative charges along the inner walls of such pores, hence electropositive ions can enter the pores more readily than can electronegative ions. Therefore the potential difference across the membrane interface ought to be predicted by the Nernst relation

$$\Delta E = RT/F \ln A_2/A_1$$

where ΔE is the potential difference, *R* and *T* the gas constant and absolute temperature, *F* the faraday; *n* the number of equivalents of ion moved, and A_1 and A_2 the chemical activities of the passively transported ion on each side of the membrane. This means that the transmembrane distribution of ions produced by the forced exchange mechanism results in the production of diffusion potentials across the negatively charged pores. According to this scheme, most of the observed potential difference is due to the high intracellular concentration of [illegible]

[illegible]

ance for this ion across the membrane. A Nernst

[illegible]

squid axon. A precise measurement of intracellular K^+ and Na^+ content and the equilibrium assumption inherent to the Nernst equation remain unsolved problems. Moreover, the postulated carrier substance has never been isolated. Nevertheless, the assumptions of forced ionic exchange with carrier-transport can explain several observations in osmoregulating systems.

An interesting example of a carrierlike concept is that of R. Goldacre. A protein molecule is pictured in the coiled state such that its charges on carboxyl and amino groups are satisfied, and such that it has no net charge. The uncharged coiled molecule diffuses to the cell membrane where it is uncoiled by an energy-requiring process and attains a net charge. It attracts ions from the medium by coulombic forces, and the ion-protein complex moves across the cell to the opposite side. There it is activated energetically to resume the coiled or spiral shape, whence its attached ions are released into the solution or ejected from inside to outside of the cell. Goldacre has used the scheme to account for osmotic work done by amebas.

Dependence on membrane properties. R. Osterhout and others have noted that NH_3 (ammonia) rather than NH_4^+ (ammonium) ions could penetrate the membrane of marine eggs and mammalian red blood cells. This deduction, made from observed staining of cells with neutral red, received support when the observation was extended to include the movement of several weak acids and bases across various biological membranes. The so-called trapping concept of diffusion of weak electrolytes was developed as follows. The *membrane is assumed* permeable for the undissociated, but *not for the* dissociated, species of a given buffer pair. Given a gradient of pH across the cell such that the pH of external fluid is less *than that in the cell,* and the molecules involved are $NH_3 - NH_4$, *then* NH_3 would diffuse down its chemical concentration gradient into *the acidic solution* where it forms NH_4OH (ammonium hydroxide) *and* dissociates forming NH_4^+. *If the pH gradient* is kept constant, NH_3 will diffuse continuously to the acid side where it accumulates NH_4^+ ions. The concentration *ratio* of $[NH_3] + [NH_4^+]$ of the fluid on each side of *the membrane* is predictable from the Henderson *relationship*. Thus it can be shown that

$$[NH_{4i}^+][OH_i^-] = [NH_{4o}^+][OH_o^-]$$

where *the subscripts denote* the concentrations inside and outside the cell.

The formation of *undissociated* molecules or of ion pairs occurs in the nonaqueous solvent phase *of a membrane*. This principle has been employed to account for the accumulation of potassium in cells. Thus K^+ *and* OH^- ions combine in the membrane to form KOH, which in turn diffuses across the membrane *toward* the cell interior, at which interface the ion pair dissociates into K^+ *and* OH^-

ions. The force causing dissociation of KOH at the cytoplasmic interface is the maintenance of a low cellular pH. This is achieved in plants by metabolic acid production, by diffusions, or by both, and by several anions in plant and in animal cells, and consequently the performance of osmotic work may be explained by the concept of trapping and ion-pair formation.

MEMBRANE PERMEABILITY

Implicit in all theories of transport of water or solutes are assumptions on the permeability properties of the membrane. For example, in a Na^+ transporting system, the carrier concept implies that NaX will diffuse rapidly across the membrane, while free Na^+ ions diffuse slowly or not at all. This is a device for potentiating diffusion of a sluggishly diffusible substance and is economical from a thermodynamic viewpoint. If the membrane were freely permeable for a transported solute, the rate of back diffusion of that solute would increase as its concentration on the transported side increased. Since concentration gradients across cell membranes are high back diffusion rates would be high Thus, the mechanism would require a large amount of energy to accomplish net transport for a solute, while overcoming the force of back diffusion of that solute. These and other considerations led to much experimental and theoretical work from which emerged concepts of membrane permeability.

Lipid-pore concept. The modern picture of membrane structure is approximately a composite of older ones of a lipid phase permeated with pores Presumably, fatty acid molecules are lined up perpendicular to the cell surface with their polar ends pointing toward the aqueous medium. Biological membranes are at least of a thickness equal to the length of two fatty acid molecules. Protein molecules, arranged so that their long axes are tangential to the membrane surface form a lacy network or mosaiclike pattern over the membrane The hydrocarbon groups link the polypeptide chains to the underlying fatty acid, while the polar groups remain in aqueous phase A protein film so adsorbed, is insoluble in water, and is about 5–50 angstroms thick and quite stable The over-all biological concept is that nonpolar molecules penetrate the membranes by dissolving in the lipoid phase, while polar substances, like the inorganic ions, can penetrate if the membrane pore diameter is larger than the ionic diameter *See* CELL MEMBRANES AND MONOLAYERS.

Resin network. A commonly used assumption pictures the pore as a cylindrical hole penetrating the membrane at a right angle to its surface The concept of the aqueous filled pore is a complicated one. G. Scatchard has postulated that the membrane is a continuous resin network formed like a sponge. The aqueous interstitial fluid forms a continuous network between the branching molecules of resin lattice. This aqueous network is functionally equivalent to a pore.

Fixed ionic charges. Assuming the presence of pores in membranes, Meyer, Sievers, and Teorell formulated the theory of fixed charges on pore walls. It is known that artificial membranes such as collodion, silicates, or protein, when interposed between two electrolyte solutions of different concentration, give rise to an electrical potential difference. In a membrane with positively charged pores, the dilute solution would be positive in an external circuit to the concentrated one; in a membrane with negatively charged pores, the reverse orienta tion would hold. According to K. Sollner, the deviation of magnitude of the membrane potential from that of free diffusion between the same two electrolytes at a liquid junction with no membrane is an inverse function of membrane porosity The assumption of fixed charges on pore walls has explained a wide variety of data in artificial perm-selective membranes. There are gaps in the theory such as the nature of Donnan forces between immovable charges and penetrating ions at the two interfaces of the membrane and within the charged pore itself. However, useful analogies between artificial perm-selective membranes and biological membranes have been made. In particular, perm-selectivity of nerve, muscle, or frog skin membranes, and diffusion potentials have been invoked to account for the electrical and osmotic proper ties of these tissues.

MECHANISMS OF WATER TRANSPORT

The mechanisms for water transport in living and in nonliving systems may be listed as follows: (1) hydrostatic forces; (2) osmotic gradients, (3) electroosmosis; and (4) miscellaneous—chem ical reactions, thermal gradients, contractile vacu ole and pinocytosis.

Hydrostatic forces. Operationally, one usually defines flow of fluid through a tube in terms of two main parameters, namely, pressure and resistance to flow It follows that a bulk flow of fluid across any boundary requires a force, and that the direc tion of such flow is down the gradient of pressure Filtration refers to the passage of water or solu tion under the influence of a hydrostatic force through the pores of a membrane interposed be tween two solutions of identical concentration A filtration process separates undissolved constituents from a solution, while ultrafiltration separates dis solved constituents (proteinate) from a solution

Gibbs-Donnan forces Ultrafiltration refers to the passage of solute and water across a membrane between two nonidentical solutions in a system like that conceived by J. Gibbs and F Donnan. The sim plest example of such a system is shown below

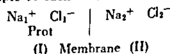
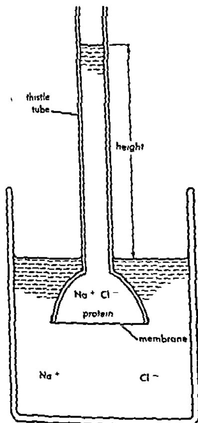


Figure 3 presents an experimental model of the system at equilibrium.

Conditions are that the membrane M permits passage of Na^+, Cl^-, and water, but not of protein



ig 2 Simple model illustrating a Gibbs-Donnan quilibrium system

te. At equilibrium, one can derive the well-known relation

$$[Na_1^+][Cl_1^-] = [Na_2^+][Cl_2^-]$$

Since $[Na_2^+] = [Cl_2^-]$ it follows that

$$[Na_1^+] + [Cl_1^-] > [Na_2^+] + [Cl_2^-]$$

This means that a hydrostatic pressure must be applied to side I to prevent the osmotic flow of solvent from II to I The osmotic conditions at equilibrium (no bulk flow) are

$$[Na_1^+] + [Cl_1^-] + [Prot] = [Na_2^+] + [Cl_2^-] + [p/RT]$$

where p is the hydrostatic pressure difference between I and II, R the gas constant, and T the absolute temperature Ultrafiltration will occur when the pressure applied to solution I is sufficient to move the ... tion o ... Mo ... tinguı ... from movement by diffusion In filtration a net movement of a finite mass of solvent occurs, while in diffusion, as it is usually defined there is no net movement of solvent By definition, the only force operative in a diffusion cell is that of the gradient of chemical potential of the transported solute material H. Ussing made use of such differences when he measured permeability coefficients of various biological membranes with and without net movement of solution.

Filtration across capillaries. Examples of hydrostatic movement of water or solution in biological systems may be found in capillary beds and in glomeruli of kidneys. As blood is pumped by the heart into the arterial side of a capillary system, the hydrostatic pressure head pushes an essentially protein-free filtrate of the blood plasma across the capillary wall into the interstitial fluid When the blood reaches the venous side of the capillaries, the blood pressure has been dissipated, and, as predicted from the balance of Donnan forces and of hydrostatic pressure forces, an osmotic flow of fluid occurs from the interstitial fluid, across the capillary wall, and into the venous capillary In the steady state, the amount of fluid filtered equals the amount of fluid reabsorbed from the capillaries This is a skeletal description of the classic concept of capillary function patterned mainly after the work of E. Starling and E. Landis Recent modifications are (1) that of R. Chambers, B Zweifach, and E Shorr, who have found evidence of capillary shunts through which blood can flow without losing or gaining protein-free fluid; and (2) that of F Chinard, who can explain data on isotopic flux in many capillary beds by invoking the action of diffusional forces. Chinard, by means of phenomenological equations, has shown that glomerular filtration must be occasioned by both osmotic and hydrostatic forces Without hydrostatic pressure, there would be no bulk flow, and theoretically, solutes would get from glomerular arteriole to glomerular capsular fluid by virtue of the diffusional forces operating across the glomerular membrane

The aforementioned considerations apply not only to all animals possessing a vascular system, but to any cellular forms containing protein-rich cytoplasm within a plasma membrane surrounded ... invertebrates and vertebrates, aquatic and terrestrial In animal cells, hydrostatic pressure is small (about 10–50 mm Hg) but sufficient to balance osmotic difference occasioned by the Donnan forces operative across the plasma membrane However, in plant cells, the hydrostatic pressure can be relatively tremendous (15–30 atmospheres) owing to the tough cellulose wall encasing the plasma membrane

Osmotic gradients. The presence of osmotic pressure differences across porous membranes is considered as one of the most important factors causing a net osmotic flow of water in biological systems The usual assumption made for the system | Solution I | Membrane | Solution II | is that the membrane is *permeable* to solvent, but not to solute Such a requirement for the membrane is approximated in highly selective cation- or anion-permeable membranes However, most living membranes such as amphibian skin and bladder, nerve,

eggs, erythrocytes, stomach, intestine, bladder, capillaries, and even mitochondria appear to be permeable to solutes of small molecular weight, that is, of molecular weights up to at least 200, as well as to water. This means that the water activity of cellular fluid tends to equilibrate rapidly with that of extracellular fluid, so that appreciable osmotic differences between the phases rarely exist. Despite wide differences in chemical composition between cellular and extracellular fluid, there exist no measurable differences of water activity. Water activity has been evaluated by the usual measures of colligative properties such as freezing point depression, melting point, or vapor tension lowering.

The *aforementioned remarks are not so general* as they might appear, because large osmotic pressure differences are present, or appear to be present, across membrane in many biological systems. All of the osmoregulating forms present such osmotic pressure differences.

Examples of systems with apparent osmotic gradients are [| body fluid | gills | pond water |] in fresh-water fish, crabs, and worms; [| body fluid | skin | pond water |] in Amphibia; [| body fluid | renal tubular cell | urine |] in kidneys producing urine either hypertonic or hypotonic to the body fluids; and [| soil water | protoplast | cytoplasm |] in plants.

Franck-Mayer hypothesis The steady-state maintenance of osmotic gradients across cells has been a biological problem for years The magnitude of some gradients (such as the urine of a hydropenic dog, 1500 milliosmoles per kilogram and plasma, 300 milliosmoles per kilogram) is too great to be explained by hydrostatic pressure gradients, except in plant cells This led to the Franck and Mayer hypothesis of osmotic gradients. Apart from maintenance of gradients, their hypothesis included a mechanism for transporting water (solvent) from a solution of high osmolarity to a solution of low osmolarity; that is, the mechanism could move water up its gradient of activity. Figure 3 illustrates the essentials of the Franck-Mayer scheme as applied to the process of formation of hypertonic

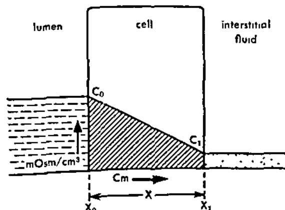


Fig 3 Scheme of an *intracellular osmotic gradient in* a kidney tubule during production of osmotically concentrated urine Osmotic activity C is plotted along the ordinate, while cell thickness X is plotted along the abscissa.

urine. When the intracellular osmotic activity C_0 at the lumen side of a renal tubular cell is slightly greater than that of the luminal fluid (urine), and the intracellular osmotic activity at the interstitial side of the cell C is equal to that of the interstitial fluid, water could be transported from the lumen to the interstitial fluid. The postulated source of solutes at C_0 was a depolymerizing reaction, enzymatically catalyzed, whence a large molecule was split into numerous small particles at the interface. Thermodynamically, the minimum free energy expenditure for maintenance of the gradient by the mechanism, regardless of its origin, must be at least equal to the heat dissipation, or to the decrease of free energy of *solute diffusing from* X_0 to X_1 The number of particles Q diffusing across an area A of cell in unit time is

$$Q = -DA\,\frac{(C_1 - C_0)}{(X_1 - X_0)}$$

for steady-state conditions. D is the diffusion coefficient $X = X_1 = X_0$, the cell thickness or path length for diffusion. The equation is an integrated form of Fick's equation with the implicit assumption of a flat sheet of cells and zero bulk flow The rate of decrease of free energy for the diffusion is

$$\frac{\partial\,\Delta F}{\partial t} = QRT\ln\,(C_0/C_1)$$

where ΔF is the change of free energy, t the time, R the gas constant, and T the absolute temperature Since diffusion is an irreversible process, the free energy loss cannot be funneled back into the transporting or gradient-creating mechanism. An evaluation of the minimum rate of expenditure of free energy, made by W. Brodsky, W. Rehm, W. Dennis and D. Miller, in the case of mammalian renal cell, yielded a value of 21,000 kcal/(kg)(hour) This is about 1000 times the maximal rate of respiration for living cells and imposes a severe limitation on the theoretical use of intracellular osmotic gradients as devices for water transport in biological systems A major limitation of the Franck-Mayer scheme is the short diffusion path, one cell width or about 20 μ. If the whole gradient were confined to the plasma membrane of water-transporting cells, the intensity of energy expenditure would be greater than that calculated, for membrane thickness is only about 100 angstroms

Countercurrent systems. An ingenious way out of the Franck-Mayer dilemma is the so-called countercurrent multiplier system of B. Hargitay, W. Kuhn and H. Wirz. This scheme was originated by analogy with the well-known physical laws of heat flow in concentric pipes containing counterflowing fluids The elements of the scheme applied to the mammalian kidney are illustrated in Fig 4 The first step in the production of osmotically concentrated urine is glomerular filtration which produces an ultrafiltrate of 300 milliosmoles/liter, isoosmotic with plasma. The filtrate in the proximal tubule suffers a reduction of volume without change of osmotic activity. Next the fluid enters the descending

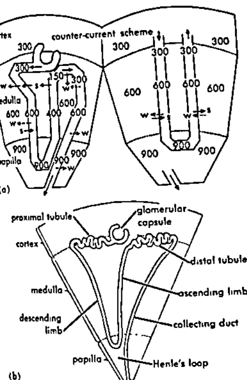


Fig 4 (a) Schematic presentation of a renal countercurrent system for production of an osmotic gradient in the interstitial fluid and consequently for production of urine of osmotic activity greater than that of plasma All numbers indicate osmotic activity in milliosmoles per liter, s, direction of the diffusion of salts, w, direction of diffusion of water (b) An anatomical reference diagram of the structures shown in a

limb of the loop of Henle where it suffers further reduction of volume, but increases in osmotic activity by equilibrating with hypertonic interstitial fluid This raises the questions of how hypertonic interstitial fluid is produced and how the osmotic ...ent from cortex to papilla is maintained.

Gottschalk has assumed that the ascending ... of the loop of Henle contains a salt pump ...h can pump salt from the lumen to the inter...al fluid and that this region of tubule is im...meable to water Such a pump would tend to ...duce a region of hypertonicity in the intersti... fluid The maintenance of the gradient under ...dy state conditions is occasioned by the countercurrent design of the vasa recta (postglomerular ...inches of arterioles from glomeruli in the sub...tical regions of the kidney). As the descending ...rt of the vasa recta plunges into the hypertonic ...pilla, salts diffuse into the blood, concentrating ...toward the hypertonic level of papillary intersti...il fluid As the ascending branch of the vasa ...nerges from the papilla, it moves salt-laden blood ...ward the cortex, a region isotonic with peripheral ...lood Therefore, salts diffuse from the blood to...ard the interstitial fluid and the osmotic gradient ... maintained

The source of energy for this process is the Na^+ ...or salt pump in the ascending limb of the loop of Henle. The lumen fluid herein is actually diluted as the tubule moves from the loop to the juxtaglomerular portion of the tubule. Between the juxtaglomerular region of the tubule and the point where it joins the collecting duct, the tubular walls are supposedly water-permeable. Water then moves from the lumen to the isotonic interstitial fluid, from which it is carried back to the body by the renal vein. This step is the true water economy mechanism for the body. A tubular fluid, isotonic with peripheral blood, then enters the cortical portion of the collecting duct. As the collecting duct proceeds toward the papilla, its contents equilibrate osmotically with those of the interstitial fluid. Since osmotic activity of interstitial fluid increases in moving from cortex to papilla, the osmotic equilibration renders the final urine hypertonic to the systemic plasma

During periods of excess water intake, the mammalian kidney elaborates a urine hypotonic to the plasma Presumably, dilution of body fluids during water ingestion reduces secretion of the antidiuretic hormone of the posterior pituitary gland Absence of this hormone may, according to Ussing, reduce the level of water permeability of the tubular cells Thus, the salt pump in the ascending limb of Henle's loop will dilute its luminal fluid, but this fluid will not equilibrate osmotically with interstitial fluid as it passes through the distal tubule and collecting duct

The biological applications of countercurrent systems have been fruitful for the explanation of diverse phenomena. It has been invoked to account for heat conservation in the extremities or skin of aquatic warm-blooded animals such as seals, whales, or ducks It may be present in gill systems, in swim bladders, or in any system containing or separating two solutions of different osmotic activity K Schmidt-Nielsen has noted the importance of diffusion length in such systems, showing that the relative length of renal papillae in desert animals is greater than that in other mammals

Osmotic gradient in plant cells. A typical plant cell is that found lining the pith, phloem, cortex, or xylem of succulent plant organs. It consists of ...

...centration to the intracellular region of relatively high salt concentration The exact mechanism for the transport could be the transmembrane transport of an anion such as nitrate (NO_3^-) at a site separate from that of cation movement Alternatively, it could be by formation of ion-pairs in the membrane In either case, the osmotic pressure difference across the membrane provides a force oriented to drive water from the exterior to the interior of the cell. Water penetrates into the cell, until the internal (turgor) pressure is sufficient to stop the flow Such turgor pressures are of the order of 10–20 atmospheres

Plasmolysis. The process (Fig 5) of shrinking the cytoplasm and vacuole by immersing an isolated

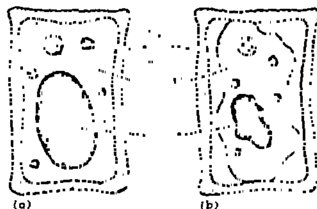


Fig. 5. Osmotic behavior of plant cell during plasmolysis. (a) The plant cell is immersed in an external medium which is hypotonic to the cytoplasm. (b) The cell is immersed in an external medium which is hypertonic to the cytoplasm.

plant cell in a solution of higher osmotic activity than that of plant cytoplasm is plasmolysis. Presumably, no osmotic shrinking or swelling would occur if the cells were immersed in a solution of the same osmotic activity as its cytoplasm. Such data on plasmolysis may not give a true measure of osmotic activity of plant cells because of solute penetration. Nevertheless, a good approximation may be made in cells from which solute leaks very slowly and into which external solute does not penetrate. The data are usually checked by cryoscopic determinations on the juice of pulverized cells. Discrepancies in the results have been found when both techniques have been applied to the same plant tissue.

Electroosmosis. If a membrane of porous clay, collodion, or silicate is placed between two identical electrolyte solutions and electrical current is passed across the whole system, a flow of solution across the membrane occurs. This phenomenon, known as electroosmosis, led to much work on artificial membranes. The implication is that an electrical force can transport water up an osmotic gradient and up a hydrostatic pressure gradient. K. Sollner has postulated that electroosmosis is related to the sign and density of fixed charges on the pore walls of membranes and the differences in mean or effective pore diameters in a given membrane. This led to the preparation of membranes of high electrochemical activity.

In the system

| Electrolyte solution I | Membrane | Electrolyte solution II |

an electric potential difference can be detected across the membrane. The magnitude and orientation of the potential depends on the nature and concentration of the two solutions, and the pore diameters and density of fixed charges on the walls of the pores. The fixed charges are negative carboxyl groups in cation-permeable membranes and positively charged amino groups ($-NH_3^+$) in anion-permeable membranes.

A perfect electronegative membrane would be permeable to cations only; hence the Nernst relation

$$E = \frac{RT}{F}\frac{a_2}{a_1}$$

would hold, implying that current through membrane pores is carried by cations only. Of biological interest are certain properties common to both living and nonliving membranes. For example, an electrical potential difference across a membrane separating two dissimilar solutions and ion selectivity in the membrane are characteristic of systems in plant cells, muscle tissue, frog skin, kidney tubules, gastric mucosa, and many others. Apart from such qualitative analogy, the fixed charge scheme of Meyer, Sievers, and Teorell and the membrane pore heterogeneity of Sollner have not been fitted precisely into a theory involving living membrane.

With externally applied current, a process known as anomalous osmosis has been observed in ion-selective membranes. Thus, when a charged membrane, such as oxyhemoglobin collodion, separates two electrolyte solutions, each of different osmotic activity, water will flow for a transient period from the concentrate to the dilute solution (negative anomalous osmosis). Under certain conditions, water flows transiently in the direction required by the osmotic force assembly, but with magnitude different from that occasioned by the osmotic force. Sollner has determined the contribution of osmotic flow due to osmotic force and due to anomalous osmosis. First, the transient flow is measured across an uncharged membrane. Then the transient flow is measured across the same membrane after it has been rendered positively or negatively charged by adjusting its pH. The difference of transient flow rates permits an estimate of ordinary and of anomalous osmosis. Since anomalous flows were detected easily in systems at physiological concentration levels, the phenomenon is of biological interest.

Miscellaneous forces related to water transport. Thermal gradients, if present in living cells, could provide a force for water movement. This is purely speculative, because the amount of water moved would be small and the direction of movement would be uncertain.

Chemical reactions with production of, or with consumption of, water would provide a device for water transport. However, the molar quantity of water involved in most known reactions is small compared to the known rates of water transport in animals.

Hydration and dehydration of protein molecules have been invoked as water-transporting processes. It is difficult to tell whether a protein molecule absorbs pure solvent or bulk solution, and it is even more difficult to see how such a system could drive water in a given direction.

Contractile vacuoles, the structures in the cytoplasm of Protozoa such as paramecia, have been observed extensively while the paramecia were im-

mersed in solutions of varying osmotic activity. The vacuole appears to fill with water, burst, and consequently extrude its contents The rate of pulsation acuole increases as the external medium is d dilute. Vacuolar activity depends on metnergy as shown by the suppressing action of . The mechanism of water *transport of con-* vacuoles could be by a secretion of water vacuole, by hydrostatic pressure forces ap- the vacuole, or by secretion of solute and n of water into the vacuole.

cytosis is a term applied to the engulfing of or solution by pseudopodlike processes of ite membranes The droplet of water is en- by the outer surface of the membrane, n it migrates to the inner surface at which t is pushed into the cell.

rohormonal control systems. Given the pres- of water- and solute transporting mechanisms, rganism requires controlling machinery to such processes so that effective osmotic reg- n will be maintained The system in mam- though complex, has been studied extensively. fore, a presentation of the elements of such a m can provide a model, applicable by analogy e *osmoregulating system of other animal* s. Anatomical elements of the mammalian sys- re the neurohypophyseal tract of the hypothal- s and posterior pituitary and the kidney tubule. ably the distal part plus collecting ducts. oreceptor cells, sensitive to small (1-4%) ges of blood osmotic activity, are present in abypophyseal tract of the hypothalamus When e cells respond to osmotic stimuli, impulses sent through the nerve tract to the posterior tion of the pituitary gland The gland responds increasing or by decreasing the rate of secretion antidiuretic hormone, ADH The hormone, after ching the blood, *is carried to the kidney where* affects the water-transporting cells of the tubule d collecting ducts so that they transport water, parently against the forces of osmosis, from the bular lumina to the blood The entire nerve-hor- one renal osmotic chain helps maintain constant molarity of the body fluids under *conditions of* cessive as well as deficient water intake

If the animal ingests a large volume of water, ual to 2-5% of its body weight, the absorbed and becomes distributed through the total body of ater and thereby dilutes the osmotic activity of ellular, interstitial, and plasma fluids by 1-3% Presumably, the osmoreceptors respond by reduc- ng their rate of transmission of impulses, and con- equently the output of ADH from the posterior pituitary to blood is reduced The absence of per- fusing hormone is supposed to reduce the osmotic permeability to water of the tubular and collecting duct cells This means that water remains in the lumen and that urine, of copious volume and hypo- tonic to blood, is produced Consequently, the in- gested load of water is eliminated, and the body fluids are concentrated back to the optimal level of osmotic activity

If the animal is deprived of water for 12-24 hours, the body fluid becomes slightly concentrated. The osmoreceptor cells transmit impulses to the posterior pituitary which in turn secretes ADH at an accelerated rate into the blood. The hormone supposedly increases the osmotic permeability of tubular and collecting duct cells Then water is removed by osmotic forces from the lumen to the hypertonic interstitial fluid surrounding the collecting duct This process concentrates the urine osmotically and reduces the volume excreted Consequently, a small amount of water without solute is returned to the body fluids. This tends to reduce the osmotic activity of body fluid toward a normal level [W.A BR.]

Bibliography: F. P. Chinard, Formation of glomerular fluid, *Am. J. Physiol.*, 171:578-606, 1952; H. T. Clarke (ed), *Ion Transport Across Membranes*, 1954; A S Crafts, H. B Currier, and C. R. Stocking, *Water in the Physiology of Plants*, 1949; A. Krogh, *Osmotic Regulation in Aquatic Animals*, 1939, C. L. Prosser (ed.), *Comparative Animal Physiology*, 1950.

Osmosis

The spontaneous flow of liquid through a natural or artificial membrane. The direction of flow is determined by the compositions of the fluids separated by the membrane The phenomenon may be observed by immersing in water a tube partially filled with an aqueous sugar solution and closed at the end with parchment An increase in the level of the liquid in the solution results from a flow of water through the parchment into the solution The process occurs as a result of a thermodynamic tendency to equalize the sugar concentrations on both sides of the barrier. The parchment permits the passage of water, but hinders that of the sugar, and is said to be semipermeable. Specially treated collodion and cellophane membranes also exhibit this behavior. These membranes are *not perfect*, and a gradual diffusion of solute molecules into the more dilute solution will occur Of all artificial membranes, a deposit of potassium ferrocyanide, $K_4Fe(CN)_6$, in the pores of a fine-grained porcelain most nearly approaches complete semipermeability

The flow of liquid through such a barrier may be stopped by applying pressure to the liquid on the side of higher solute concentration The applied pressure required to prevent the flow of solvent across a perfectly semipermeable membrane is called the osmotic pressure and is a characteristic of the solution.

The walls of cells in living organisms permit the passage of water and certain solutes, while preventing the passage of other solutes, usually of relatively high molecular weight. These walls act as selectively permeable membranes, and allow osmosis to occur between the interior of the cell and the surrounding media See EDEMA, PLANT, WATER RELATIONS OF, SOLUTION [F-J J]

Osteichthyes

The bony fishes, one of three classes of Recent fishlike vertebrates which includes most of the familiar fishes. The class is similar to the Chondrichthyes or cartilaginous fishes, and contrasts with the Agnatha in having jaws, paired nostrils, true teeth, paired fins and girdles (unless lost secondarily), three semicircular canals, and scales (unless lost or modified). The Osteichthyes contrast with the Chondrichthyes in having a bony skeleton, a swimbladder (at least primitively), a true gill cover, and mesodermal ganoid, cycloid, or ctenoid scales. These scales are sometimes absent or modified. Fertilization is usually external, but if it is internal, the intromittent organ is not derived from pelvic-fin claspers. Most often a modified anal fin or a fleshy tube or sheath functions in sperm transfer.

Phylogeny. It is now believed that bony fishes preceded cartilaginous fishes. This interpretation runs counter to the long-accepted belief that cartilage preceded bone in phylogeny as well as in ontogeny. This assumption received support from the largely cartilaginous endoskeleton of lungfishes and sturgeons, which are living representatives of ancient groups. It was also believed that sharks were ancestral to bony fishes. It is noted, however, that the oldest fishes were bony, and that in such temporal lineages as the old actinopterygian fishes and the lungfishes, ancient forms were better ossified than are their recent derivatives. Thus cartilage may be regarded as having been acquired secondarily. The adaptive advantage of cartilage in developmental stages appears to explain satisfactorily its prevalence in young fishes.

Classification. The Osteichthyes are divided into two subclasses, the Actinopterygii, or ray-fin fishes, and the Sarcopterygii, or choanate fishes The latter are subdivided into the two superorders, Crossopterygii and Dipnoi It is instructive to note that bony fishes made their first appearance in the Middle Devonian, with the more or less simultaneous appearance of the Actinopterygii, the Crossopterygii, and the Dipnoi. The three groups were well differentiated at that time; it seems clear, then, that the history of the group extends further back than the record thus far reveals.

The Sarcopterygii were destined to experience moderate success until early Mesozoic time after which they barely persist. Today's fauna includes only three families, four genera, and seven species of this subclass. The major contribution of the group has been its successful paternity, in the crossopterygian line, of amphibians and higher classes of vertebrates In the late Paleozoic the Actinopterygii, developing somewhat more slowly, realized marked success in the palaeonisciform line, a group that was replaced by ascendant development of amiiforms and related groups in the Mesozoic, to be succeeded by the enormous outbursts of clupeiform, cypriniform, and perciform successors in the Upper Cretaceous and Eocene. Most modern fishes belong to these same groups which are conveniently termed teleosts. The Recent fauna of the class Osteichthyes includes 2 subclasses, 31 orders, about 333 families, very roughly 3,100 genera, and probably between 15,000 and 17,000 species.

A classification scheme of the Osteichthyes follows. Equivalent names of each group are given in parentheses. See separate articles on each group

Class Osteichthyes
- Subclass Actinopterygii
  - Order Polypteriformes (Cladistia)
  - Order Acipenseriformes (Chondrostei)
  - Order Semionotiformes (Protospondyli and Ginglymodi)
  - Order Amiiformes (Halecomorphi)
  - Order Clupeiformes (Isospondyli and Haplomi)
  - Order Myctophiformes (Iniomi)
  - Order Saccopharyngiformes (Lyomeri)
  - Order Cypriniformes (Ostariophysi)
  - Order Anguilliformes (Apodes)
  - Order Notacanthiformes (Heteromi)
  - [illegible]
  - [illegible]
  - [illegible] Solenichthyes)
  - Order Gadiformes (Anacanthini)
  - Order Lampridiformes (Allotriognathi)
  - Order Percopsiformes (Salmopercae)
  - Order Beryciformes (Berycomorphi)
  - Order Zeiformes (Zeomorphi)
  - Order Perciformes (Acanthopterygii)
  - Order Pegasiformes (Hypostomides)
  - Order Pleuronectiformes (Heterosomata)
  - Order Echeneiformes (Discocephali)
  - Order Tetraodontiformes (Plectognathi)
  - Order Gobiesociformes (Xenopterygii)
  - Order Batrachoidiformes (Haplodoci)
  - Order Lophiiformes (Pediculati)
  - Order Mastacembeliformes (Opisthomi)
  - Order Synbranchiformes (Symbranchii)
- Subclass Sarcopterygii (Amphibioidei)
  - Superorder Crossopterygii
    - Order Osteolepiformes (Rhipidistia)
    - Order Coelacanthiformes (Coelacanthini and Actinistia)
  - Superorder Dipnoi (Dipneusta)

Currently, there is a lack of general agreement as to the nomenclature of higher groups of fishes. Broadly, two systems are employed, that of C T Regan and the more recent one of L. Berg In the Berg system, ordinal names, like family names, are formed by the addition of a standard suffix, -formes, to the stem of a type genus. See ACTINOPTERYGII; OSTEICHTHYES FOSSILS; PISCES (ZOOLOGY); SARCOPTERYGII. [R M B]

Bibliography: L. S. Berg, *Classification of Fishes, Both Recent and Fossil*, 1947.

Osteichthyes fossils

The class Osteichthyes includes all the higher bony fishes as distinguished from the more primitive ostracoderms and placod[illegible] h also had a bony

olutionary changes in ray finned fishes

ructure	Chondrostean (palaeoniscoid)	Holostean	Primitive teleost	Advanced teleost
kull roof	Nearly smooth	Nearly smooth	With crests and depressions	With crests and depressions
Cheek area	Covered with bones	Covered with bones	Open	Open
Notochord	Rodlike	Still present, replaced in some by bony vertebrae	Replaced by bony vertebrae	Possesses vertebrae
Dorsal fin	Single	Single	Single	Spiny-rayed part, soft-rayed part
Caudal fin	Notochord extends into upper lobe	Notochord ends at base of fin, upper lobe composed of rays only	Rays attached to enlarged bones at end of vertebral column	Rays attached to enlarged bones at end of vertebral column
Upper jaw	Fastened to cheek bones	Freed from cheek	Attached only at snout	Attached only at snout
Paired fins	Broad bases, scales on front borders	Narrow bases	Narrow bases; in primitive position	Modified in position: pelvic, forward, pectoral, elevated
Anal fin	Scales along front border	Scales along front border	No scales on front border	Spines plus rays

skeleton, and from the sharks which have a cartilaginous skeleton (Fig 1). The osteichthyan bony skeleton is a retention, with modifications and elaborations, of that in the ancestral placoderms The skull, in particular, has a basic architecture which was early modified into three types· the actinopterygian, the crossopterygian, and the dipnoan *See* OSTRACODERMI; PLACODERMI

Osteichthyan classification (A S Romer, 1945; L S Berg, 1958) is unsettled mainly because the phyletic relationships of the included groups are poorly understood. A conservative arrangement is therefore followed here

Class Osteichthyes
- Subclass Actinopterygii, ray-finned fishes
  - Superorder Chondrostei, palaeoniscoids, sturgeons
  - Superorder Holostei, semionotoids, amioids
  - Superorder Teleostei, clupeoids
- Subclass Sarcopterygii (= Choanichthyes), lobe finned fishes
  - Order Crossopterygii, rhipidistians and coelacanths
  - Order Dipnoi, lungfishes

The placoderm ancestry of the Osteichthyes is unknown With the first appearance of the class in the Devonian, it was already separated into the actinopterygians and the sarcopterygians

Subclass Actinopterygii. In addition to a characteristic skull pattern, the actinopterygians may be readily distinguished by the structure of the paired fins which have the rays attached to a small internal skeleton at the fin base. The three superorders Chondrostei, Holostei, and Teleostei are actually evolutionary grades (see table) rather than natural ...

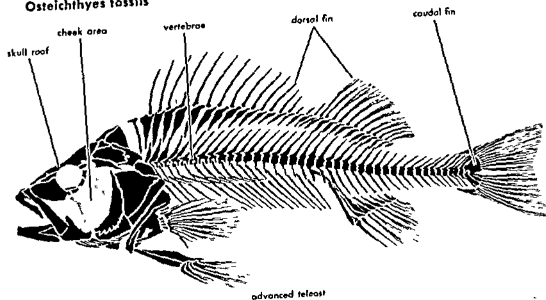
te...

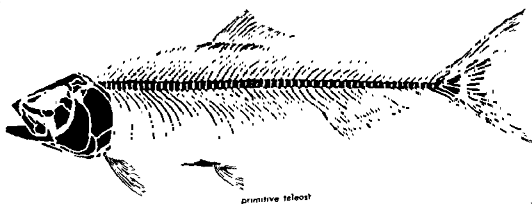
... from the Devonian to the Jurassic. They gave rise to a more progressive group, the sub-holosteans (mostly of the Triassic), which combined palaeoniscoid and holostean characters in various ways, and were transitional to the Holostei The transition included changes in the feeding mechanism, in the opercular region, in the braincase, in the fins, and in the tail.

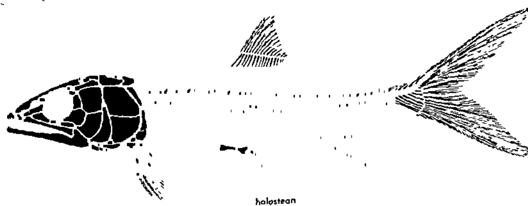
Superorder Holostei There are at least four major evolutionary lines at the holostean level: (1) the semionotoids, which first appeared in the late Permian and include the living gars; (2) the more diversified amioids which began in the Triassic and include the modern *Amia*, (3) the long-snouted aspidorhynchoids which lasted from the late Jurassic through the Cretaceous; and (4) the Mesozoic pholidophoroids which gave rise to the teleosts The aberrant pycnodonts, frequently regarded as holosteans, may have arisen directly from the palaeoniscoids.

Superorder Teleostei. The pholidophoroid-teleost transition probably occurred in the late Triassic and involved no major morphologic change The lowest teleost level, represented by the heterogeneous isospondyl assemblage, retained abdominal pelvic fins and other pholidophoroid characters which were lost in the more advanced groups. The ...

Cretaceous They introduced the highest level of teleost organization, the mostly marine Acanthopterygii, with spines on the fins, thoracic pelvic fins, and a characteristic modification of the hyoid apparatus Most modern teleost families which have a fossil record appear abruptly at the beginning of the Tertiary. This can only mean that there was an explosive radiation during the Cretaceous in both fresh water and marine environments ...







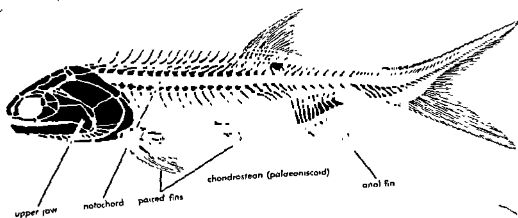


Fig. 1. The ray-finned fishes

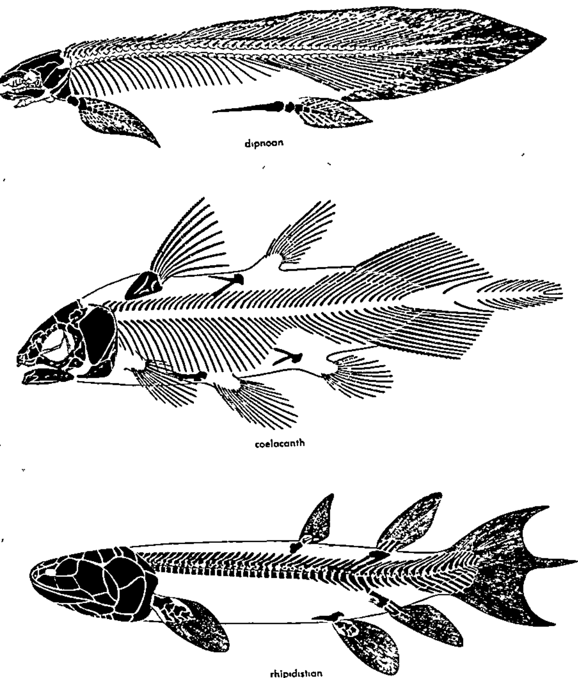


Fig 2 The lobe-finned fishes

are poorly represented in the fossil record The ostariophysids, percoids, scombroids, blennioids, plectognaths and scorpaenoids are examples of groups which were fully differentiated in the Eocene

Subclass Sarcopterygii. The term Choanichthyes is no longer appropriate for this subclass because the coelacanths, and possibly the Dipnoi, do not have choanae. The class Sarcopterygii (substitute name suggested by A. S Romer, 1955) was sharply separated into the Crossopterygii and Dipnoi at least as early as the Early Devonian The structure of the paired fins, with an internal skeleton of ...erygial type, the structure of the scales and the presence of two dorsal fins indicate, among other resemblances, that they probably had a common origin and that they are only distantly related to the actinopterygians (Fig. 2).

Order Crossopterygii The crossopterygians were divided into two branches in the Devonian, the rhipidistians and the coelacanths. The fresh-water rhipidistians, which were heavy-bodied, predaceous fishes, gave rise to the tetrapods, also in the Devonian This is demonstrated by the structure of the skull, the vertebrae, and the skeleton of the paired fins The rhipidistians became extinct during the Permian The coelacanths evolved from primitive rhipidistians and developed distinctive

modifications in the skull and the post-cranial skeleton. During the course of their long history, they inhabited both fresh-water and marine environments. Coelacanths have not been found in Tertiary rocks, but the living *Latimeria* closely resembles its Mesozoic ancestors.

Order Dipnoi. The dipnoans, with their crushing dentition and characteristic skull structure, have been a strictly fresh-water group since the Devonian. Like the coelacanths, they have had a conservative history, mostly because of a reduction in ossification and modifications in the fins and scales [B.S.]

Bibliography: L. S. Berg, "System der Rezenten und Fossilen fischartigen und Fische," Veb Deutscher Verlag der Wissenschaften, Berlin, 1958, (translated from 1955 Russian edition); A. S. Romer, *Vertebrate Paleontology*, 2d ed., 1945.

Osteolepiformes

An order of fossil crossopterygians also known as the Rhipidistia and Osteolepidoti. *Osteolepis macrolepidotus* is the best-known species The body of *Osteolepis* was covered with cosmoid scales A pair of small premaxillae and large maxillae were present and armed with small teeth. The tail was heterocercal and the paired fins were bluntly lobed. A pineal "eye" occurred in this species. *See* CROSSOPTERYGII. [C.B.C.]

Osteomyelitis

Inflammation of a bone and its adjacent bone marrow, usually resulting from infection. Pyogenic (pus-forming) organisms, such as the hemolytic staphylococci, are the most frequent agents in acute cases. Chronic forms occur and may be induced by tuberculosis, syphilis, and other microorganisms. All forms have decreased in incidence since the advent of the antibiotics *See* STAPHYLOCOCCUS, SYPHILIS; TUBERCULOSIS.

Pyogenic infections may be blood-borne in a series of conditions which vary from relatively mild local infections to fully developed bacteremias The simple act of chewing, for example, may cause the release of bacteria from periodontal abscesses into the blood. Other routes of infection include local or direct spread from a neighboring soft-tissue lesion such as a boil or abscess, or may be due to contamination of bone exposed during trauma.

The long tubular bones of the arms and legs are most often involved Bone destruction can become extensive if prompt diagnosis and treatment is not initiated The acute phase is marked by suppuration, that is, a pus-forming inflammatory reaction A persistent osteomyelitis tends to become chronic and shows combined reparative attempts mingled with continued tissue breakdown. The acute clinical course is marked by fever, chills, leukocytosis, and local symptoms of inflammation such as pain, swelling, and redness of the affected part.

Chronic osteomyelitis may show little direct evidence of its presence, and attention may be directed to it only incidentally or as the result of fracture of a weakened bone after minor stress In other cases, however, a chronic osteomyelitis may give constant signs and symptoms of its presence, depending on the etiologic agent and the bone involved *See* BONE. [E.C.S.T.]

Osteostraci

An extinct order of Agnatha or jawless vertebrates, known from the Late Silurian and Devonian of Europe, Asia, and North America. Also known as Cephalaspida, they were mostly small, ranging from about 2 in. to 2 ft in length. The head and part of the body were encased in a solid armor of bone, while the posterior part of the body and the tail were covered with thick scales Some early forms lacked paired fins, though most possessed flaplike pectoral fins. One or two dorsal fins were present. Their depressed shape and the position of the eyes on the top of the head suggest that Osteostraci were bottom dwellers (Fig. 1). The throat region

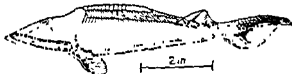


Fig. 1. The ostracoderm, *Hemicyclaspis*, a cephalaspid, a very early jawless vertebrate of the Lower Devonian. About one-half natural size. (From E H Colbert, *Evolution of the Vertebrates*, Wiley, 1955)

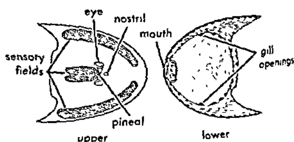


Fig 2 The upper and lower surfaces of the head in the Devonian ostracoderm, *Cephalaspis*, a cephalaspid. (From E. H. Colbert, *Evolution of the Vertebrates*, Wiley, 1955)

was covered below by small plates and had a small mouth in front. The large gill chamber contained as many as 10 pairs of gills, each opening separately to the exterior (Fig 2). It is believed that they fed by sucking in water and straining out the contained food particles in the gills Because the internal skeleton of the head was often bony, its structure is well known. Details of this structure—particularly the single, dorsal, median nostril—and the presence of only two pairs of semicircular canals indicate a relationship to extinct Anaspida and to some living Cyclostomata. All are grouped in the subclass Cephalaspidomorphi. The Devonian *Cephalaspis* is the best known genus. *See* AGNATHA; ANASPIDA; CEPHALASPIDOMORPHI; CYCLOSTOMATA (CHORDATA); OSTRACODERM. [R.H.DE.]

stracoda

a order of the class Crustacea considered by some thorities to be a subclass. They are bivalved or-nisms, world-wide in distribution, found in a vari-

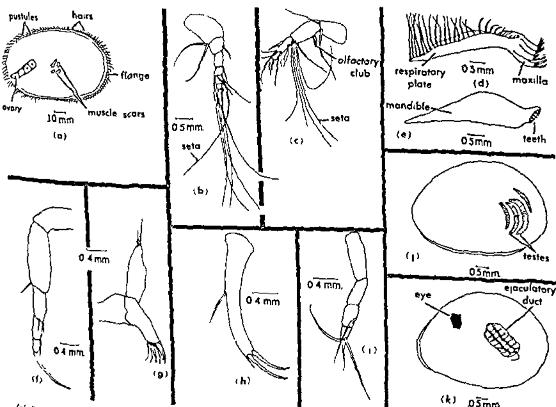
ving species of ostracods are free-living; how-ver, members of the genus *Entocythere* live as com-mensals on the gills of crayfish The reported spe-es of the genus *Sphaeromicola* live commensally n amphipods and isopods *Cypridopsis hartwigi* Müller is predaceous, preying upon snails of the pecies *Bullinus contortus* and *Planorbis glabratus*. See AMPHIPODA; ISOPODA.

LIVING OSTRACODS

Extant ostracods are divided into the Myodocopa, Platycopa, Cladocopa, and Podocopa In addition to these, paleontologists add the Paleocopa, which are extinct. These five groups are classified as orders or suborders by different authorities See separate articles on each group.

Morphology. Knowledge of the morphology of ostracods has been obtained mainly from studies of representatives of the fresh-water family Cypridae The ostracod shell is composed of three layers: a thin waxy outer layer, a hard mineralized portion consisting of an organic mesh impregnated with calcium carbonate and other minerals, and an inner chitinous coating composed principally of chitin and proteins The valves range in length from 0 35 mm, in the fresh water species *Cypridopsis yucatanensis* Furtos, to several inches in some species. The shell is commonly pigmented, and may be covered with spines, tubercles, hairs, pustules, and pore canals. The valves are closed by the action of adductor muscles.

Appendages Ostracods typically possess seven pairs of segmented appendages and one unsegmented pair. Usually, there are four pairs of postoral appendages; however, species of the suborder Platycopa have just three pairs, and the Cladocopa have two pairs of appendages posterior to the mouth. The appendages starting from the anterior are first antennae, second antennae, mandibles, maxillae, first thoracic legs, second thoracic legs, third thoracic legs, and the *unsegmented caudal rami*, or *furcae*. The thoracic appendages, though commonly referred to as walking legs, are not always adapted to locomotion.



(a) Right valve of *Cypricercus tuberculatus* (Sharpe 1908). (b) *Cypria turneri* Hoff 1942, first antenna (c) *Cypria turneri*, second antenna (d) *Cypricercus tuberculatus* (Sharpe 1908), maxilla (e) *Cypricercus tuberculatus*, mandible. (f) *Candona hoffi* Ferguson 1953, first thoracic appendage of female. (g) *Candona hoffi*, second thoracic appendage of female. (h) *Candona hoffi*, third thoracic appendage of female (i) *Candona hoffi*, caudal ramus (furca) of female (j) *Cypria ophthalmica* (Jurine 1820), mesial view of right valve showing testes. (k) *Physocypria pustulosa* (Sharpe 1897), lateral view of the left valve of male

Digestive system. The alimentary canal of ostracods consists of the atrium, or mouth, the esophagus, stomach, intestine, hindgut, and anus. The hepatopancreases, or livers, are paired glands situated in the hypodermis; they discharge their secretions into the stomach through ducts.

Excretory system. The excretory system of the fresh-water Cypridae includes a pair of excretory glands (glands of the first antennae) located alongside the esophagus, a pair of shell glands, and a single maxillary gland. Each of these glands possesses a dorsal urinary canal and a ventral endsac.

Respiratory system. Among the Ostracoda, the typical respiratory structures, common in most orders of Crustacea, are absent. Gills have been observed in only a few species of the subfamily Asteropinae. The respiratory epithelium of the inner lamellae of the shell, the branchial plates of the appendages, and the natatory setae of the antennae are considered among the more important respiratory organs of ostracods.

Nervous system. The central portion of the nervous system consists of a cerebrum, a circumesophageal ganglion, and a ventral series of fused ganglia. The principal nerves are the optic, antennulary, antennary, labral, labial, mandibular, maxillary, thoracic (legs), and abdominal.

The sense organs include a conspicuous, heavily pigmented eye, absent in the Cladocopa and Platycopa, that is situated in the anterodorsal portion of the body Auditory structures, in the form of ellipsoidal bodies, are located at the bases of the first antennae. An olfactory club is attached to the antepenultimate podomere (segment of an appendage) of the second antenna Sensory hairs are present on both pairs of antennae and on the distal podomeres of the mandibular palpi.

Muscles In addition to the adductor muscles, well developed muscles are attached to both pairs of antennae and to the labium, labrum, mandibles, maxillae, thoracic legs, and furcae.

Reproductive organs The reproductive organs follow the bilaterality common to other organs of the body. Ostracods are dioecious, that is, the sexes are separate. The female reproductive system includes the ovaries in the posteroventral region of the shell, the oviducts, seminal receptacles, and vaginae. The male organs of generation are the testes, that appear as circuitous bandlike organs; the vasa deferentia, coiled threadlike structures resembling a coiled watch spring; the ductus ejaculatorius which is placed posterodorsally; and the paired penis, situated posterior to the ejaculatory duct.

Production and Development. Reproduction is always sexual. Both parthenogenetic (development of an unfertilized egg) and syngamic (development of a fertilized egg) types of reproduction are common among members of this order. Some species, for example, *Cypridopsis vidua* (Muller), reproduce only by parthenogenesis. In other species, syngamy apparently is the sole method of reproduction. Many species are oviparous (egg-laying), while others, such as *Darwinula stevensoni* (Brady and Robertson), are viviparous (bear living young)

The spermatozoa are long, threadlike structures, frequently longer than the adult ostracod The eggs are usually round or oval in shape Many species produce eggs that are brightly colored Due to its calcium-impregnated, double-wall shell and the presence of a large quantity of water within its cavity, the ostracod egg is able to remain viable during long periods of drought. Because the ostracods have the ability to withstand desiccation, one investigator was able to culture them in aquaria from mud imported to Norway from South America, Africa, and Australia.

The egg of the fresh-water Cypridae hatches into a bivalved nauplius. The nauplius has three pairs of appendages, the first and second antennae and the mandibles. During the course of development, nine instars appear, the last being the sexually mature adult [E.F]

FOSSIL OSTRACODS

These have been studied intensively since they were found to be useful in petroleum geology Certain species and genera lived only during a relatively brief geologic interval, and are important as guide fossils, or keys to geologic age (see INDEX FOSSIL). Like other microfossils, their small size permits the paleontologist to obtain complete faunas from well cores. Ostracods were numerous in the past, and occur in many kinds of sediments

Classification is difficult because the soft parts by which living ostracods are classified, are only very rarely fossilized It is therefore based on shape of the carapace, nature of the hinge, the way one valve overlaps the other, muscle scars, the presence of dimorphism, lobation, and presence or absence of such structures as duplicature, frill, eyespot or brood pouch Many families, superfamilies, and suborders are extinct, and possessed structures not known in living ostracods. These forms have no living descendants to which they can be compared Other difficulties stem from incomplete descriptions and inadequate illustrations in the older literature Immature instars (developmental stages) must be recognized as such, and included in the same species as the adults

Orientation of extinct ostracods is done by comparison with living forms. In general, adductor muscles migrate forward during ontogeny (development of the individual) and are anterior in the adult The posterior part of the carapace is the widest, and the anterior part is highest. From these criteria, the relative positions of other structures can be determined

The range of fossil Ostracoda is questionable Certain flexible, partly calcified carapaces from Cambrian strata are thought by some to be Ostracoda, if so, they are the oldest known Undoubted ostracods occur from the Ordovician to the Recent

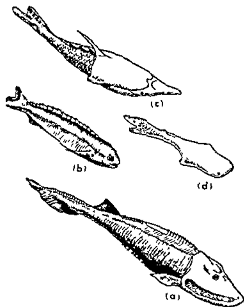
Many fossil ostracods belong to the suborders Podocopa and Myodocopa, which are still extant Others, from Paleozoic rocks, are classified as the

order Paleocopa. *See* CLADOCOPA; CRUSTACEA; ...DOCOPA; PALEOCOPA; PLATYCOPA; PODOCOPA; *also* PARTHENOCENESIS. [R.V.K.]

Bibliography: R S. Bassler and B Kellett, *Bibliographic Index of Paleozoic Ostracoda*, Geol Soc. Spec Papers 1, 1934; E Ferguson, Jr., A new ...nd Ostracod from Maryland, *J Wash Acad* 43 194-197, 1953; C. C. Hoff, *The Ostracods of ...ois*, Illinois Biol Monograph, 19.1-196, 1942; V Kesling, *The Morphology of Ostracod Molt ...ges*, Illinois Biol. Monograph, 21:1-324, 1951. W Müller, Ostracoda, *Das Tierreich* 31·1-434. ...2

...tracoderm

...opular name applied to several groups of ex...t jawless vertebrates (fishes). Most of them ...e covered with an external skeleton or armor of ...e, from which is derived their name, meaning ...ell skinned" They are known from the Ordovi...n Silurian, and Devonian periods, and thus in...de the earliest known vertebrates Together with ...ir presumed descendants, the living Cyclosto...ta, they form the class Agnatha *See* AGNATHA, ...CLOSTOMATA (CHORDATA)

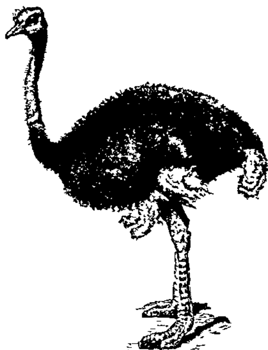


...stracoderms, jawless vertebrates of Silurian and De...nian age, drawn to the same scale (a) *Hemicyclas...s* was a cephalaspid or Osteostraci (b) *Pterolepis*, ...anaspid (c) *Pteraspis*, a pteraspid or Heterostraci ...*Thelodus*, a coelolepid (From E H Colbert, *Evo...tion of the Vertebrates*, Wiley, 1955)

The following groups are generally considered ...ostracoderms: (1) Osteostraci, bottom-living ...rms with the head and anterior part of the body ...cased in a single shield, (2) Anaspida relatives ...f Osteostraci with a more normal fishlike shape ...nd with small, thick scales; (3) Heterostraci, a ...oup unrelated to Osteostraci and Anaspida with ...he head and anterior part of the body enclosed in an armor formed of several or numerous plates; (4) Coelolepida, small forms whose skin is set with minute scales *See* ANASPIDA; COELOLEPIDA; HETEROSTRACI; OSTEOSTRACI. [R H DE.]

Ostrich

Any of four large, terrestrial, flightless birds of the genus *Struthio*, order Struthioniformes, which are widely distributed in Africa. Best known is the



The ostrich, *Struthio* sp, height to 8 ft. (From P. M Duncan, ed, *Cassell's Natural History*, Cassell)

ostrich of northern Africa, *S camelus*. This is the largest of all living birds; the males attain a height of 8 ft and weigh 200 lb Ostrich eggs and flesh are used as food The birds are reared in captivity as novelties and for their plumes which are plucked from living birds The plumes are widely used in millinery and for other ornamental purposes

Somewhat similar to the ostrich, but structurally different and classified in separate orders, are the emus of Australia, the rheas of South America, and the now extinct moas of New Zealand *See* STRUTHIONIFORMES [J.D B.]

Otter

Either of two aquatic or semiaquatic carnivores of the family Mustelidae, one is marine, the other a fresh-water species The fur of both is valuable; for example at one time that of the sea otter, *Enhydra lutris*, commanded prices of $2500 or more apiece The sea otter is now present only in very limited numbers in the Aleutian Islands and off southern California, and is completely protected The river otter, *Lutra canadensis*, formerly ranged



The otter, *Lutra canadensis;* length 4–5 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

over all except the arid parts of North America, but is now extinct throughout much of its original range. However, it is still moderately common in the southern swamps, where a few pelts are harvested each year. It is an intelligent, playful animal, attaining a length of 4 ft and weighing up to 20 lb. The sea otter is about the same length but much heavier-bodied; a large adult may weigh 80 lb. *See* CARNIVORA. [J.D.B.]

Otto cycle

Spark ignition reciprocating engines such as are used in automobiles operate in accordance with the Otto cycle N. A. Otto (1832–1891) built a highly successful engine that used the sequence of engine operations proposed by Beau de Rochas in 1862 The Otto cycle can be adapted to run either as a four-stroke engine cycle or as a two-stroke engine cycle

Four-stroke cycle. The engine mechanism consists of a piston in a cylinder. An intake valve and an exhaust valve are located in the top or head end of the cylinder (Fig 1).

In operation a charge, consisting of an explosive mixture of air and fuel, is drawn into the engine cylinder through the intake valve as the piston moves down from its initial top center position. At the end of this suction stroke, the intake valve closes, and the piston moves back on its second or compression stroke. Shortly before the piston has reached the top of its stroke, a spark ignites the fuel mixture, and the resulting combustion rapidly raises the temperature and pressure of the gas held by the top center position of the piston On the third stroke (downward), expansion of the gases does useful work. Near the bottom of this expansion stroke, the exhaust valve opens and some exhaust gases rush out. The bulk of the remaining exhaust gases are pushed out by the sweep of the piston on the fourth (upward) return stroke of the cycle. The exhaust valve then closes, the intake valve opens, and the cycle is repeated. In this engine arrangement, one power stroke is obtained for every four strokes of the piston, or for every two revolutions of the crankshaft.

Two-stroke cycle. The two-stroke cycle engine requires an auxiliary air pump, which introduces the new charge by blowing the spent gases out of the cylinder. In this way, it is possible to obtain one power stroke for every revolution of the crankshaft. This two-stroke spark-ignition arrangement is less efficient than the four-stroke engine because extra air is handled beyond that required for combustion, and because some incoming fresh charge is lost in scavenging the spent gases from the cylinder between cycles.

Actual engine process. The actual processes of an internal combustion engine depart widely from the ideal cycle. The actual cycle uses a mixture of air and a complex chemical fuel which is either a volatile liquid or a gas. The rate of the combustion process and the intermediate steps through which it proceeds must be established The combustion process shifts the analysis from one set of chemicals, constituting the incoming mixture, to a new set representing the burned products of combustion. Determination of temperatures and pressures

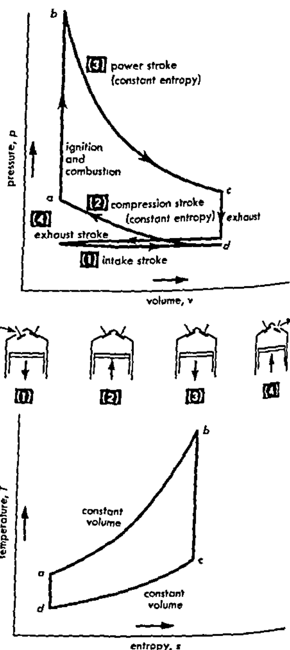


Fig 1. Air-standard Otto cycle.

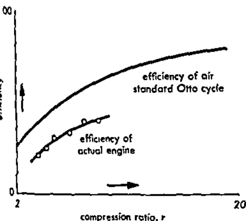


Fig 2 Effect of compression ratio on efficiency of air-standard Otto cycle and on actual Otto engine

at each point of the periodic sequence of processes requires information on such factors as variable specific heats, dissociation, chemical equilibrium, and heat transfer to and from th...

...to use the air standard cycle (*see* THERMODYNAMIC CYCLE) to analyze the Otto cycle Such a procedure substitutes a thermodynamic cycle for an engine performing with changing mass and a periodic repetition of the processes The employment of an air-standard cycle gives considerable insight into the characteristics of important noncyclic gas power plants, and also paves the way for possible future advances in which cyclic gas power plants may play a role in conjunction with nuclear heat sources

The ...
tity o...
frictı...
insulated and uncovered for heat transfer purposes With the piston at top center, the compressed air is initially exposed to heat transfer from a hot body This constant volume heat addition process, marked *a b* on Fig 1, is followed by insulation of the cylinder head and an isentropic expansion *b-c* to the bottom of the piston travel Then the cylinder head is provided with a cold body so that the path *c-d* represents a constant-volume heat rejection process Insulation is returned to the cylinder head for the last process *d a*, which is an isentropic compression ...

sta...
aul...
for ... actual Otto engine, but the trends are similar as illustrated by Fig 2 This graph shows that efficiency increases with the compression ratio, but the situation becomes one of diminishing returns at the higher-compression ratios. The efficiency of both the air standard and the actual Otto engine increases with increasing compression ratio

The efficiency of the idealized air-standard Otto cycle can be expressed by

$$\eta = 1 - \left(\frac{1}{r_v}\right)^{k-1}$$

η = efficiency
r_v = volumetric compression ratio
k = ratio of the specific heat at constant pressure to the specific heat at constant volume, it has the value of 1.40 for air

In 1958, laboratory evidence indicated that actual Otto engines have peak efficiencies at a compression ratio of about 17:1. Above this ratio, efficiency falls The most probable explanation offered is that the extreme peak pressures associated with high compression ratios cause increasing amounts of dissociation of the combustion products. This dissociation near the beginning of the expansion stroke exerts a more deleterious effect on efficiency than the corresponding gain due to increasing the compression ratio beyond 17:1. [J.B.]

Ovary

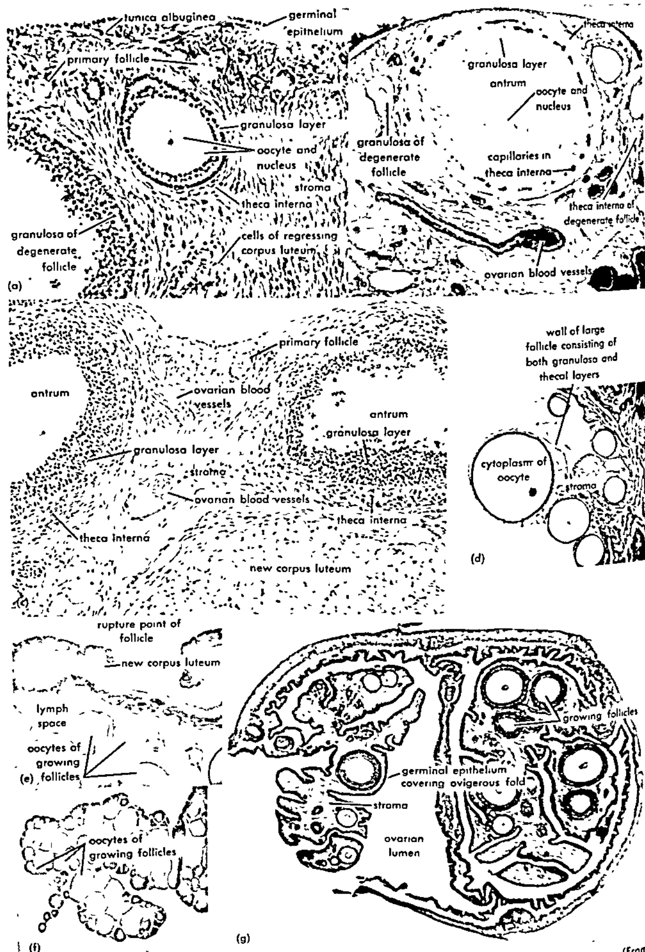
A part of the reproductive system of all female vertebrates. Although not vital to individual survival, the ovary is vital to perpetuation of the species The function of the ovary is to produce the female germ cells or ova, and, in some species, to elaborate hormones that assist in regulating the reproductive cycle

Although the ovaries develop as bilateral structures in all vertebrates, adult asymmetry is found in certain species of all vertebrates from the elasmobranchs to the mammals This asymmetry may be morphological as a result of fusion or atrophy, or it may be physiological.

The position of the ovaries in the coelom varies within different vertebrate groups Those animals producing large numbers of eggs possess ovaries that almost fill the coelomic cavity during the breeding season, but most vertebrates have relatively small ovaries The ovaries may be anchored to the dorsal body wall anywhere between the transverse septum (lower vertebrates) and the pelvic cavity (higher mammals) *See* GONAD

Histology. The ovary of all vertebrates functions in essentially the same manner. However, the ovarian histology of the various groups differs considerably Even such a fundamental element as the ovum exhibits differences in various groups The ovum of oviparous forms is large, and it synthesizes and stores large amounts of yolk. The ovum of viviparous forms, especially the mammals, is small and contains little yolk.

Mammalian ovary. The mammalian ovary is attached to the dorsal body wall by a mesovarium The free surface of the ovary is covered by a modified peritoneum called the germinal epithelium, shown in illustration (*a*). This layer may consist of a single or a stratified layer of cells, and the cell shape may vary from squamous to columnar, depending upon the species. The activity of this layer, especially during the embryonic development of the gonad whereby cells are proliferated into the interior, is responsible for its name The



(a) Ovary of pika, *Ochotona princeps* (b) Ovary of pocket gopher. (c) Ovary of prairie dog, *Cynomys leucurus* (d) Ovary of chicken, *Gallus domesticus* (From W. Andrew, *Textbook of Comparative Histology*, Oxford, 1959) (e) Ovary of lizard, *Xantusia vigilis* (Courtesy Dr M Miller) (f) Ovary of frog, *Rana* sp (From W Andrew, *Textbook of Comparative Histology*, Oxford, 1959) (g) Ovary of viviparous fish, *Neotoca bilineata* (From G Mendoza, *Biol Bull*, vol. 84, 1943)

potentialities of this layer are still a matter of controversy

Just beneath the germinal epithelium is a layer of fibrous connective tissue, varying in thickness and density in different species and at different phases of the reproductive cycle in the same species. This is the tunica albuginea, shown in illustration (*a*). Most of the rest of the ovary is made up of a more cellular and more loosely arranged connective tissue (stroma) in which are embedded the germinal, endocrine, vascular, and nervous elements (*a,b*).

The most obvious ovarian structures are the fol- [illegible]

[illegible] entiation of the perifollicular stroma to form a fibrocellular envelope called the theca interna (*b,c*) Finally, a fluid-filled antrum develops in the granulosa layer, resulting in a vesicular follicle (*b*).

The cells of the theca interna hypertrophy during follicular growth and many capillaries invade the layer (*b*), thus forming the endocrine element that is thought to secrete estrogen. The other known endocrine structure is the corpus luteum (*c*) which is primarily the product of hypertrophy of the granulosa cells remaining after the follicular wall ruptures to release the ovum Ingrowths of connective tissue from the theca interna deliver capillaries to vascularize the hypertrophied follicle cells of this newly formed corpus luteum; the hormone progesterone is secreted here

Most follicles degenerate before attaining ovulatory size (*a,b*, and *c*) In some species the theca interna of such follicles hypertrophies to form the so-called interstitial cells (*b*) The function of this tissue is unknown.

Avian ovary The avian ovary has much the same general arrangement as that of the mammal the chief difference being the size of the mature ova Other differences are a greatly reduced amount of stroma, the thinness of the theca interna, the reduced number of follicle-cell layers, and the absence of follicular antra and of corpora lutea (*d*)

Reptilian and amphibian ovary The reptilian ovary is quite similar in appearance to that of the bird The follicles are similar, and some viviparous forms are reported to have corpora lutea, although the physiology of them is uncertain Centrally placed, epithelial lined spaces said to be lymphatic spaces, are present in reptilian ovaries (*e*)

The amphibian ovary is similar to the reptilian type but contains many more follicles which are of reduced size. No corpora lutea are formed (*f*)

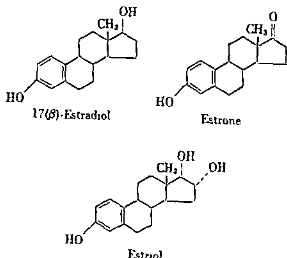
Fishes The fishes represent a diverse group as far as ovarian morphology is concerned. Many forms have an arrangement similar to the amphibians. Others have the germinal epithelium covering numerous folds within a centrally placed ovarian lumen instead of covering the outer surface (*g*)

Some elasmobranchs develop corpora lutea, but the endocrine function is doubtful. [K.L.D.]

Physiology. The ovaries function to produce both the female gametes (ova) and hormones There are about 400,000 primordial follicles in the ovaries of the human infant, but this number decreases progressively throughout life and after the menopause they may be absent. During the sexual life of the human female the number of eggs released from the ovaries does not ordinarily exceed about 400, this means that large numbers of follicles degenerate (atresia) without ovulating After ovulation, the cavity of the ruptured follicle fills with large cells containing a yellow pigment; this constitutes an important endocrine structure called the corpus luteum. Cyclic changes in the female tract are regulated by the ovarian hormones which in turn are conditioned by the pituitary gonadotrophins See OVUM; PITUITARY GLAND

Ovarian hormones Ovarian hormones fall into two classes: (1) estrogens which produce sexual receptivity (heat), vaginal cornification, and other changes in the genital tissues when administered to spayed subjects, and (2) progesterone, the hormone of the corpus luteum, which cooperates with estrogen to produce full mammary development and to prepare the uterus for implantation of the blastocyst Progesterone also functions in the maintenance of pregnancy.

Formulas of the principal natural estrogens are



17(β)-Estradiol

Estrone

Estriol

Estrogenic compounds have been identified in lower animals and in a great variety of plants. Under certain conditions, they can produce tumors of the reproductive organs as well as at other sites Cancer of the mammary glands in mice has been produced by 17(β)-estradiol, estrone, equilin, and equilenin

17(β)-Estradiol is thought to be the major hormone synthesized by the maturing ovarian follicle. Although the ovary and placenta are the main sites of estrogen synthesis, some estrogen is produced by the adrenal cortex and by the testis It is generally believed that estriol is the end product of es-

trogen metabolism, according to the equation

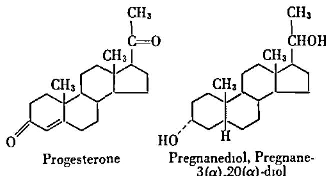


The interconversion of estrone and 17(β)-estradiol can be accomplished by microorganisms of the genus *Pseudomonas*. The human placenta contains a dehydrogenase which catalyzes the interconversion of estrone and 17(β)-estradiol, but the enzyme system does not affect 17(α)-estradiol, an estrogen present in the urine of mares.

Testosterone and progesterone probably serve as intermediates in the biosynthesis of estrogens. Placental tissue can form estrone from testosterone by way of Δ^4-androstene-3,17-dione; ovarian tissue also forms estrogens from testosterone and converts progesterone to Δ^4-androstene-3,17-dione. Estrogen catabolism and interconversion occur chiefly in the liver, the metabolites being eliminated through the bile and urine.

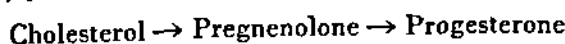
A number of unnatural estrogens have been prepared synthetically, the most important of which is stilbestrol. This compound is highly potent, can be prepared cheaply, and is active when administered orally.

In man and certain other mammals, pregnanediol is a major urinary metabolite of progesterone:



Progesterone   Pregnanediol, Pregnane-3(α),20(α)-diol

Progesterone is elaborated in relatively large amounts by the corpus luteum, and during pregnancy, by the placenta. Some is present in the adrenal cortex, where it serves as an intermediate in the biosynthesis of the adrenocortical hormones. In these various tissues the main biosynthetic pathway probably is



In the liver progesterone is reduced to pregnanediol, conjugated with glucuronic acid, and is excreted in the urine as a biologically inactive glucuronide. See ESTROGEN.

Menstrual cycle. The human menstrual cycle is about 30 days in length and is divided into several phases on the basis of endometrial histology. During the preovulatory stage one or more follicles begin to grow and the secretion of estrogen increases; the endometrium thickens but the uterine glands remain relatively straight and unbranched. Ovulation, occurring near the middle of the cycle, is followed by the progestational stage. Progesterone, acting upon an estrogen-primed endometrium, causes the stroma to become loose and edematous and the glands to secrete actively. This is the only kind of endometrium in which a blastocyst can implant. If pregnancy does not ensue, the endometrium again degenerates and is partly sloughed at the next mensis.

Complete development of the uterus requires both estrogen from the follicles and progesterone from the corpus luteum. Anovulatory cycles are superficially indistinguishable from ovulatory cycles, but if a corpus luteum is not formed, the progestational endometrium fails to differentiate.

The rhythmic nature of ovarian activity is due to a reciprocal relationship between the gonads and the anterior hypophysis (pituitary). The three gonadotrophins in proper combinations cause follicular growth, secretion of estrogen, ovulation, formation of corpora lutea, and secretion of progesterone. The ovarian hormones in turn regulate the release of pituitary gonadotrophins. Cyclic changes in the female are abolished by ovariectomy or by hypophysectomy. See REPRODUCTIVE SYSTEM.

Estrus. Estrus cycles are characteristic of subhuman mammals, in which it is only at estrus that the female is receptive to the male. The estrus periods may recur throughout the year or may be limited to a particular season. In laboratory rats and mice, estrus recurs every 4–5 days; ovulation takes place spontaneously at this time. There are many modifications of the estrus cycle in different species. The guinea pig's cycle lasts for about 16 days; in rabbits, the estrus periods are prolonged and ovulation depends upon copulation or some comparable stimulation. The dog comes into heat twice a year, the intervening quiescent period being termed anestrus.

Endocrine mechanisms that operate during the estrus cycle of the rat have been carefully studied. The follicular stimulating hormone (FSH) seems to be released at a rather uniform rate throughout the cycle, and its primary action is to stimulate growth of the ovarian follicles. The luteinizing hormone (LH) acts synergistically with FSH and promotes the secretion of estrogen. Larger amounts of LH synergize with FSH to produce ovulation. It appears that LH and luteotrophin act together to cause growth of the corpus luteum and, with a relative increase in the titers of luteotrophin, progesterone secretion begins. Estrogen is the chief hormone secreted by the ovary before ovulation; the secretion of progesterone begins shortly before ovulation and increases markedly thereafter. During the normal cycle the release of LH from the anterior pituitary is probably conditioned by the circulating levels of progesterone and estrogen. When the corpora lutea begin to wane the titers of ovarian hormones are reduced and, as a consequence, a new cycle is initiated under the influence of FSH. During pregnancy the corpora lutea persist for longer periods than usual; in certain species, new corpora lutea may be formed. See ESTRUS.

Placenta. The placenta performs an endocrine function during pregnancy and seems to serve as an adjunct to the ovaries and the anterior hypo-

physis In man and monkeys, the ovaries may be removed after the first month of pregnancy without terminating pregnancy Certain species may be hypophysectomized after the first half of gestation with normal continuation of pregnancy. It is generally believed that the placenta secretes estrogens, progesterone, and gonadotrophin. There is some evidence that it also produces luteotrophin and growth hormone. *See* PLACENTATION.

Lactation Lactation is a complicated process requiring complete development of the mammary glands, an adequate supply of milk precursors, activation of the alveolar secretory cells, and evacuation of milk Hormones are essential in all these processes Although estrogen seems to be responsible chiefly for growth of the mammary ducts, a combination of estrogen and progesterone produces the best alveolar development. In certain species, the ejection of milk is accomplished by the reflexive release of oxytocin from the posterior pituitary body. A large number of hormones undoubtedly influence milk secretion, but the site and mode of action remain obscure. In addition to luteotrophin it is probable that growth hormone, adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), thyroxin, insulin, and the parathyroid hormone all directly or indirectly influence milk secretion *See* LACTATION, MAMMARY GLAND [C D T]

Bibliography: *See* REPRODUCTIVE SYSTEM

Ovenbird

A terrestrial American wood warbler, *Seiurus aurocapillus*, breeds from Canada and Montana southward into Georgia and Arkansas This warbler is a thrushlike bird and lives on the forest floor, generally where there is underbrush The nest, built on the ground, is covered, with the entrance at the side; the bird derives its name from the shape of the nest. The ovenbird is olive-green above and white below, with black streaks on the sides, breast, and throat. It has a dark orange crown, bordered with black; its legs are pink. The ovenbird is shy, and is usually heard rather than seen. Its ringing "teacher, teacher, teacher" song, each note louder than the one before, is one of the most memorable sounds of the spring forest. *See* PASSERIFORMES. [J.D B]



The ovenbird, *Seiurus aurocapillus*, length to 6½ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Overdrive

An automotive device supplied as special equipment and containing a step-up planetary gear arrangement located between the transmission and propeller shaft Overdrive permits the propeller shaft to be driven at transmission output shaft speed, or faster than transmission output shaft speed when the overdrive comes into operation.

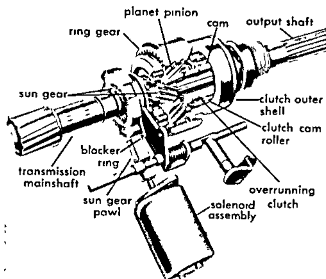
Purpose. For a given forward speed of the car, overdrive gives lower engine speed, quieter operation, and reduced gasoline consumption Most overdrives used in the United States reduce engine speed 28% for a given car speed For example, without overdrive action, a car engine might turn at 2000 rpm to achieve a speed of 40 mph. As overdrive comes into operation, engine speed could drop to about 1400 rpm at the same car speed of 40 mph This provides quieter operation and reduced gasoline consumption and engine wear.

Operation. The overdrive contains two major components, a planetary gear system and a freewheeling device (or overrunning clutch). *See* CLUTCH, PLANETARY GEAR TRAIN.

The overrunning clutch has two functions. It locks the transmission mainshaft and the overdrive output shaft together when the planetary gear system is inactive, providing direct drive through the overdrive The second function is to permit the output shaft to overrun the transmission mainshaft when the planetary gear train is in action, providing overdrive The overrunning clutch consists of an outer shell attached to the output shaft as illustrated. A circular inner member attached to the transmission shaft carries a series of cams or flats; hardened steel rollers ride in each cam. When the clutch is in action, the rollers wedge between the cams and outer shell so that the cams drive the outer shell. During overrunning, the outer shell rotates faster than the cams, the rollers move from loaded contact with the cams to disengage the clutch and permit overrunning.

The planetary gear system contains a sun gear which floats freely on the transmission mainshaft in direct drive. In overdrive, the sun gear is held stationary by a pawl that enters one of the slots in the sun gear plate which is splined to the sun gear.

With the sun gear stationary, the planet pinions are forced to rotate as they are carried around the sun gear by rotation of the planet pinion cage splined to the transmission mainshaft Pinion teeth meshed with the sun gear are momentarily at rest. The teeth opposite, which are meshed with the ring gear, are therefore required to move faster than the



Cutaway view of overdrive mechanism

cage. The ring gear is thus driven faster than the cage and transmission mainshaft. The ring gear is integral with the clutch outer shell and the output shaft Therefore, with the sun gear stationary, the output shaft is driven faster than the mainshaft.

Overdrive controls. The overdrive will not operate at low car speed However, when operating speed is reached (around 20 mph in many cars), a governor that senses car speed connects the solenoid to the car battery. The solenoid plunger compresses an internal spring which urges the pawl against a ledge on the blocker ring. The blocker ring is frictionally mounted to the sun gear plate, but it does not rotate. It has only limited freedom of rotary movement.

The driver of the car must make a conscious movement to cause the overdrive to go into operation. This movement is a momentary release of the accelerator pedal. As this takes place, the engine slows down and causes a reversal of rotation of the sun gear. As the sun gear reverses rotation, friction on the blocker ring causes it to turn a few degrees. The ledge on the blocker ring moves out from under the pawl, allowing the pawl to drop into one of the notches in the sun gear plate, locking the plate, and thus the sun gear With the sun gear stationary, the overdrive goes into action

When the car speed drops below an established speed, the governor deenergizes the solenoid, permitting withdrawal of the pawl to restore direct drive. When direct drive is desired at higher speeds for additional passing ability, depressing the accelerator pedal past the full-throttle position will interrupt the ignition for a fraction of a second, causing an instantaneous reversal of the sun gear and permitting return to direct drive by the process just described.

In reverse gear, the lockout ring is moved rearward, its internal splines connecting the external splines on the member attached to the transmission output shaft to the overdrive output shaft. Use of the friction torque of the engine in first or second transmission gear for hill descent is accomplished by locking out the overrunning clutch by pulling out a knob under the dashboard

Automatic transmissions, with torque ratios considerably greater than those of standard layshaft transmissions, have permitted the use of rear-axle ratios equivalent to the over-all ratios previously used by overdrives This fact, together with cost considerations, has resulted in application of the overdrive to standard layshaft transmissions only The automatic transmissions generally have a means of obtaining increased passing ability by depression of the accelerator pedal past the full throttle position, similar to that on overdrives *See* TRANSMISSION, AUTOMOTIVE. [F R M]

Overtone (music and acoustics)

An upper partial tone. The word overtone appeared in English about a century ago as an unfortunate translation of the German word *Oberton* As a consequence of the distinction "upper," a sixth overtone, for example, is the same as the seventh partial tone, so that numbering tends to become confused The word overtone has been used as a synonym for partial, harmonic, or mode of vibration, and the word tone has still other meanings in music. It seems desirable, therefore, not to use overtone at all, but to select one of the other words that has a more specific meaning. *See* HARMONIC (PERIODIC PHENOMENA); MODE OF VIBRATION; PARTIAL TONE; TONE (MUSIC AND ACOUSTICS) [R W.Y]

Overvoltage

The difference between electrode potential under electrolysis conditions and the thermodynamic value of the electrode potential in the absence of electrolysis for the same experimental conditions, sometimes called overpotential It is expressed in volts and is often quoted in absolute value *See* ELECTRODE POTENTIAL.

Overvoltage phenomena were first studied for electrolytic evolution of hydrogen and oxygen Understanding of overvoltage phenomena is essential in industrial electrochemistry since they are often a major factor in determining the efficiency of electrode processes *See* DECOMPOSITION POTENTIAL, ELECTROCHEMICAL PROCESS; ELECTROLYSIS.

The overvoltage for a given electrode reaction depends on electrolysis conditions such as current density (current per unit area), concentration of electrolyzed substance, temperature, presence of foreign electrolytes, nature of the solvent, and absorption of foreign substances on the electrode The overvoltage for given electrolysis conditions varies greatly from one electrolyzed substance to another—from less than 1 millivolt to a few volts

Overvoltage values are determined from the variations of electrode potential with current density by the following method The electrolytic cell is connected to a direct-current power supply and the potential of the electrode being studied is measured against a reference electrode for different current densities. The ohmic drop, or voltage drop, in solu-

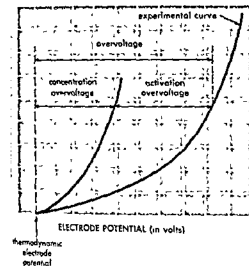


Fig 1 Current-potential curve

tion must be eliminated or greatly minimized in such measurements; and special methods have been developed to achieve this result Results are plotted as a *current-potential* curve, and the overvoltage is read on this curve.

Concentration overvoltage and activation overvoltage are the two components into which the overvoltage can generally be *decomposed*

Concentration overvoltage. This type of overvoltage arises from local variations of the concentration of reactants or products of electrolysis, or both near the electrode For instance, in the electrolytic deposition of a metal from an aqueous solution of one of its salts, the *concentration* of metal ions being consumed at the electrode is smaller near the electrode than in the bulk of solution This is called concentration *polarization* As a result, the thermodynamic potential calculated for the concentrations at the electrode surface under electrolysis conditions is different from the potential in the absence of electrolysis The difference between these two potentials is the concentration overvoltage

The consumption of reactants and consequently the concentration overvoltage increases with current density for given electrolysis conditions Conversely, concentration polarization for a given current density is [illegible] the [illegible] ingl [illegible] late [illegible] the transport of substances to and from the electrode can be treated mathematically Detailed theory has been developed for transport by convection and diffusion, for application to analytical chemistry *see* POLAROGRAPHIC ANALYSIS.

Activation overvoltage. The other component of overvoltage results from the relative slowness of electrode reactions An energy barrier must be overcome by the substances involved in an electrode reaction just as for purely chemical reactions Electrical energy in addition to that required by the thermodynamics of the electrode reaction is expended to overcome this energy barrier. Activation overvoltage corresponds to this additional consumption of electrical energy

The activation overvoltage increases with the velocity of the electrode reaction, that is, with current density. For many processes and for overvoltages exceeding 0 1 volt in absolute value, the current density is essentially an exponential function of activation overvoltage over a fairly wide range of current densities (perhaps several orders of magnitude). This relationship follows the general exponential dependence of a reaction velocity on an energy term This term for electrode processes is proportional to the activation overvoltage

It follows from the relationship between current density and activation overvoltage that a plot of activation overvoltage against logarithm of current density is linear for overvoltages exceeding approximately 0 1 volt in absolute value (Tafel law). Several Tafel lines with different slopes are obtained when the reaction mechanism changes from one range of current densities to another.

The Tafel law was established experimentally. Theoretical justification has been developed and has shown the approximate nature of this law. More rigorous analysis requires consideration of the electrolyte structure near the electrode.

The Tafel law does not hold for small overvoltages because of the effect of the backward electrode reaction Thus, an electrode process with a single rate-determining step can be conceived as the result of two opposite processes, the forward and the backward reactions. The *net velocity* and the corresponding *experimental current* density is the algebraic sum of the current densities for the forward and backward electrode reactions. Each current density term varies exponentially with activation overvoltage, and the net current density is

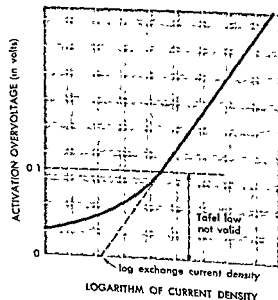


Fig 2 Tafel plot

the sum of two exponential functions of activation overvoltage. One of these exponential functions becomes negligible and the Tafel law is obeyed when the activation overvoltage exceeds 0.1 volt in absolute value

In the absence of electrolysis, the net current is equal to zero, and the current densities for the forward and backward electrode reactions are the same in absolute value but of opposite sign. The exchange current density defined in the absence of electrolysis for the forward and backward reactions cannot be measured directly but can be obtained by extrapolation of the Tafel line to zero activation overvoltage. The exchange current density is a characteristic of an electrode reaction because it is independent of current density, whereas this is not the case for activation overvoltage. However, electrolysis conditions such as temperature, nature of solvent, and presence of foreign electrolyte affect the exchange current density for a given electrode reaction Exchange current densities are generally expressed in units amp/cm^2 and, in general, for unit concentrations of reactants.

The larger the exchange current density, the inherently faster is the electrode reaction. The activation overvoltage required for different processes for identical conditions decreases as the exchange current density increases. When the exchange current density is so large that the activation overvoltage is negligible under given electrolysis conditions, the electrode process is said to be reversible for these conditions. Irreversible electrode processes have a nonnegligible activation overvoltage. The terms reversible and irreversible, although not rigorous, are convenient and in general use.

Since only concentration overvoltage can be determined for reversible electrode processes under usual electrolysis conditions, the exchange current density for such processes is obtained by special methods of nonsteady state electrolysis The effect of concentration polarization is minimized in these methods by rapid variation of one of the electrical variables, electrode potential or current, and by measurement of the resulting change of the other electrical variable. The exchange current density is determined from oscillographic recordings of current-time curves (potentiostatic method) or potential-time curves (galvanostatic method) Duration of electrolysis in these methods may vary from a few microseconds to perhaps 1 second. Another method for the study of reversible electrode processes involves periodic variations of electrode potential and current (electrolysis with superimposed alternating current).

Overvoltage phenomena are sometimes complicated by coupling of the electrode reaction with a purely chemical reaction or a crystallization process (metal deposition). The chemical reaction may precede or follow the electrochemical step. For instance, the activation overvoltage for hydrogen evolution on certain metals depends on the rate of formation of hydrogen molecules by recombination of pairs of hydrogen atoms. Experimental methods for the study of such effects and theoretical analysis have been developed. [P.D]

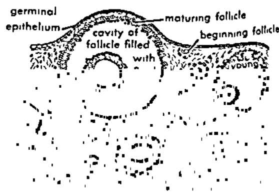
Bibliography: G. Kortum, *Lehrbuch der Elektrochemie*, 1957; E. C. Potter, *Electrochemistry*, 1956.

Ovum

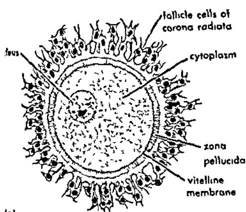
The term that designates the egg or female sex cell While strictly speaking it refers to this cell when it is ready for fertilization, the term is often applied to earlier or later stages Confusion is avoided by the use of qualifying adjectives such as immature, ripe, mature, fertilized, or developing ova The mature ova are generally of spheroidal shape and of large size. In fact, the largest known cells of a living animal are represented by the mature ova of the ostrich and the shark *Chlamydoselache*, which are about 8 cm in diameter. Among oviparous animals, which spawn eggs at or before the time of fertilization, those that produce larvae capable of feeding at an early stage have small eggs, and their development is generally characterized by radical transitions in appearance, called metamorphosis, before the adult form is attained. The typically viviparous mammals, in which the developing embryo receives nourishment for growth through the uterine tissues of the mother, also characteristically have relatively small eggs which are about 0 1 mm in diameter The number of ova produced at one time varies in different animals, from millions in many marine animals that spawn into the surrounding sea water, to about a dozen or less in mammals, in which adaptations for internal nourishment of the developing embryo and care of the young are highly developed

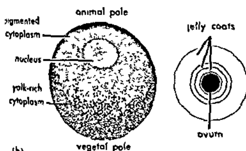
In the ovary, the immature ovum is associated with follicle cells through which it receives material for growth In mammals, as the egg matures, these cells arrange themselves into a structure known as the Graafian, or vesicular, follicle, consisting of a large fluid-filled cavity into which the ovum, surrounded by several layers of cells, projects from the layer of follicle cells that constitutes the inner wall. The fluid contains estrogenic female sex hormone

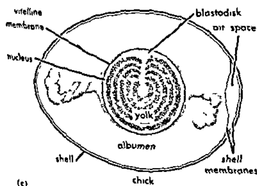
Yolk, or deutoplasm, is essentially a food reserve in the form of small spherules, present to a greater or lesser extent in all eggs. It accounts

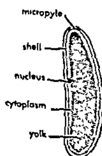


Section of a mammalian ovary.









Representative types of ova. (a) Isolecithal, human (b) Telolecithal, frog (c) Telolecithal, hen (d) Centrolecithal, fly

largely for the differences in egg size. Eggs are classified according to the distribution of yolk. In the isolecithal type there is a nearly uniform distribution through the cytoplasm, as in most small eggs. The yolk in telolecithal eggs is increasingly concentrated toward one pole, as in the large eggs of fish, amphibians, reptiles, and birds Centrolecithal, or centrally located yolk, occurs in eggs of insects and cephalopod mollusks.

Polarity of organization is manifest, in telolecithal eggs, by the higher concentration of yolk at the vegetal than at the animal pole, the distribution being radially symmetrical about the polar axis This distribution is present, though less marked, in isolecithal eggs The animal pole of the egg is also the place where the polar bodies are formed (see GAMETOGENESIS; OOGENESIS). It marks the region where the future head of the embryo will develop.

Bilateral symmetry is sometimes evident in the shape of the egg, as in insects. The bisecting plane represents the future plane of bilateral symmetry of the embryo. In chickens, the plane of bilateral symmetry of the developing embryo is generally perpendicular to the long axis of the shell, and its left side is towards the blunt end of the shell. In frogs and many invertebrates, it has been discovered that the point at which the fertilizing spermatozoon enters the egg determines the position of the future plane of bilateral symmetry of the embryo, including its dorsoventral axis. The significance of this may be appreciated when it is realized that, together with the polar axis, the position of the dorsoventral axis specifies completely the future location of all the organs of the developing embryo and adult The establishment of these axes provides the basic pattern of internal organization of the egg *See* ANIMAL SYMMETRY.

Regulative and mosaic eggs. Despite this patterning at an early stage, it has been possible experimentally, in many species of animals, to obtain complete individuals from fragments of unfertilized or of developing eggs (see EMBRYOLOGY, EXPERIMENTAL) This is generally successful when the cuts are made through the polar axis Eggs in which such fragments develop as a whole are termed regulative and are widely distributed throughout the animal kingdom, including man, as evidenced by the occurrence of identical, multiple births Eggs in which the fragments develop as structurally defective embryos have been termed mosaic and have been reported principally among the annelids, mollusks, and tunicates. However, the distinction has lost much of its original significance in recent years with the demonstration that twins and double monsters were also experimentally obtainable in eggs of the latter group by appropriate procedures. Experiments involving cutting across the polar axis have provided a further clue as to the nature of the internal organization of the egg When thus cut in half, each fragment develops defectively, but a combination of animal and vegetal quarters develops into a normal indi-

vidual, as does also the remaining middle fragment. Thus, there are interactions of materials distributed along the polar axis, and a proper balance is essential for normal development

Egg membranes. The membranes which surround the egg are designated as primary, secondary, and tertiary according to their derivation from the ovum itself, the surrounding follicle cells of the ovary, and the lining of the oviduct, respectively. The primary membrane, in practically all animals, is also termed a vitelline membrane. In many animals this becomes elevated from the surface upon fertilization and is termed a fertilization membrane Secondary membranes are represented by the zona pellucida and surrounding gelatinous, follicle cell-containing material known as the cumulus oophorus in the mammalian egg Often, however, it is difficult to decide if a particular coat of the egg is produced by the egg itself, by the surrounding follicle cells, or by both. The distinction also loses some of its significance in view of the evidence that most of the materials of the growing oocyte are supplied to it in practically fully formed state. Tertiary membranes are illustrated by the gelatinous coats of amphibian eggs and by the white, or albumen, shell membranes and the leathery or calcareous shells of eggs of reptiles and birds These structures are applied to the egg as it descends the oviduct after ovulation

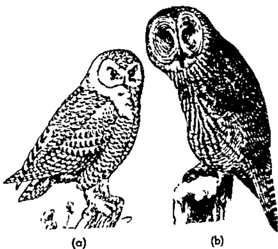
In higher plants, the phanerogams, the egg nucleus of the embryo sac represents the equivalent of the ovum of animals. In lower plants, the cryptogams, egg formation in an archegonium is more nearly analogous to the production of eggs in the ovary in animals. [A TY.]

Owl

Any member of the order Strigiformes, a group of predatory birds of cosmopolitan distribution There are 2 living families and 143 species. Owls are nocturnal birds with soft plumage and a fringe on the leading edge of each wing which enables them to fly noiselessly. They have strong, hooked beaks and strong talons. The large, round head is frequently equipped with ear tufts. The eyes are large, directed forward, and specially adapted for nocturnal activities. The young are downy when they hatch

Owls feed primarily upon mammals, but they also eat birds, other land vertebrates, and sometimes insects and other arthropods. The preferred food for most of them appears to be rodents. Owls swallow their prey whole or in large pieces The undigestible bones, fur, and feathers are rolled into balls, called pellets, and disgorged from the mouth

The family Tytonidae, the barn owls, is a small family with only 10 species, but is almost cosmopolitan in distribution. The barn owl, *Tyto alba*, absent only in New Zealand and Hawaii, is the only barn owl in the United States. This bird, sometimes called the monkey-faced owl, is unusually shy and secretive and may live for years in a barn or hollow tree without being noticed.



Owl. (a) Snowy, *Nyctea scandiaca*, length to 27 in (b) Great gray, *Strix nebulosa*; length to 33 in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

All of the other owls in the United States belong to the family Strigidae. This is also a cosmopolitan family, but the species are not so far-ranging as the barn owl. Of the 133 species in this family, 17 occur in the United States The great horned owl, *Bubo virginianus*, is fairly common throughout North and South America. It is almost 2 ft long and has prominent ear tufts, the only large owl with these tufts. Its color varies sharply in different localities The somewhat smaller barred owl, *Strix varia*, shares with the great horned owl the name of hoot owl in many localities. The barred owl is barred with dark and light gray markings, sometimes brownish or buff. It does not have ear tufts In the United States this owl is found in the deciduous forests east of the Great Plains.

The screech owl, *Otus asio*, a small owl with short ear tufts, is equally familiar both by sight and by its quavering cry. It is found throughout most of North America There are two color phases, not related to sex or locality, one predominantly gray, the other rufous; sometimes both hatch in the same nest. This little owl, only 10 in. long, feeds upon insects, mice, small birds, and other animals *See* STRIGIFORMES. [J.D B]

Oxalate

A salt or ester of oxalic acid, HOOCCOOH, in which one or both acidic hydrogens have been replaced by a metal or an organic radical One or both of the carboxylic hydrogens can be replaced to produce an extensive series of salts, esters, and mixed (double) salts In general, alkali-metal salts are water-soluble, others are insoluble. Calcium and potassium hydrogen oxalate are found in many plants, the former is also found in urinary calculi Many salts have practical applications, for example, in pyrotechnics (sodium), blue printing (potassium-iron), and analytical procedures (ammonium). Esters of the simpler alcohols hydrolyze readily; others are more stable. Diethyl oxalate is used as a solvent, as a dyestu intermediate, and in plastics. *See* ESTER; O [F.H H.]

Oxalic acid

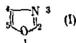
A white solid with melting point 189.5°C (decomposition) that is obtainable by careful drying of oxalic acid dihydrate (melting point 101.5°C). Oxalic acid, $(COOH)_2$, is the first of the series of dicarboxylic acids. Its salts (oxalates) are prevalent in nature, for example, KHC_2O_4 in plants of the oxalis family (wood sorrel), and CaC_2O_4 in eucalyptus bark.

Sodium oxalate forms when sawdust is fused with sodium hydroxide, and together with hydrogen on heating sodium formate in vacuum at 300°C (commercial process). Nitric acid oxidation of sucrose or starch gives the acid directly.

The acid is liberated from its salts by addition of dilute sulfuric acid. It is used as a bleaching agent for rust and ink stains, in textile and leather production and as monoglyceryl oxalate in the manufacture of allyl alcohol and formic acid. Easily oxidized it is determined by titration with $KMnO_4$ in dilute sulfuric acid. When heated with concentrated H_2SO_4, it gives equal volumes of CO and CO_2. *See* CARBOXYLIC ACID; OXALATE. [E B B.]

Oxazole

One of a group of organic heterocyclic compounds containing oxygen and nitrogen in the 1 and 3 positions of a five-membered diunsaturated ring. Formula (I) shows the structure and numbering system for a typical member of the group, 1,3-oxazole



See AZOLE; HETEROCYCLIC COMPOUNDS

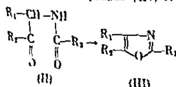
Properties. The parent compound (I) is a colorless, volatile, weakly basic liquid (bp 69–70°C), with an odor resembling that of pyridine. Oxazole is miscible with water and organic solvents. Mineral acids form salts that tend to dissociate in water. Ox...

...ance to disr...

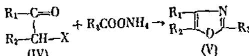
The nucleus ...

position being the usual point of attack. Hydrogenation over a platinum catalyst or with sodium and alcohol gives tetrahydro derivatives (oxazolidines) or ring-cleavage products.

Preparation. Standard syntheses start with α-acylamidocarbonyl compounds (II), or with α-

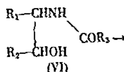


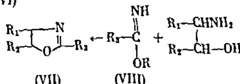
haloketones (IV). Cyclization of (II) with sulfuric acid or phosphorus pentachloride generates an oxazole (III). The R groups may be aryl or alkyl. The reaction of α-haloketones (IV) with ammonium salts of carboxylic acids gives oxazoles (V). Carboxy oxazoles obtained from these syntheses can be decarboxylated with relative ease, and



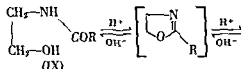
therefore serve as useful intermediates in syntheses of carboxyl-free oxazoles.

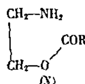
2-Oxazolines (VII) are formed by cyclization of





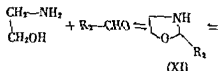
β-hydroxyalkylamides (VI) or by reaction of β-aminoalcohols with iminoethers (VIII). Hot aqueous acid hydrolyzes 2-oxazolines to *O*-acylethanolamines or to ethanolamines. Alkali converts 2-oxazolines to *N*-acylethanolamines or to ethanolamines. 2-Oxazolines are intermediates in the acid- or base-catalyzed interconversion of *O*-acyl- (X)

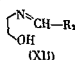




and *N*-acylethanolamines (IX).

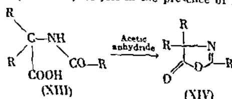
Tetrahydrooxazoles, or oxazolidines (XI), are formed from ethanolamines and aldehydes. The process can be reversed. Oxazolidines under suit-





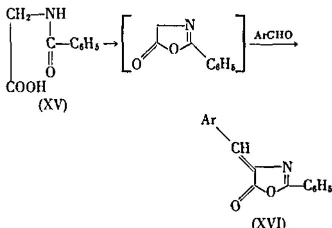
able conditions exist either in equilibrium with, or entirely in the form of, the isomeric imine (XII).

Azlactones, or 5-oxo-2-oxazolines (XIV), are generally formed by cyclization of α-acylamido acids (XIII). Hydrolysis in the presence of acids



or bases regenerates the original α-acylamido acid. Alcohols or amines react to give the corresponding

α-acylamido ester or amide, respectively. When an aldehyde, generally aromatic, is warmed with acetic anhydride and *N*-benzoylglycine, hippuric acid (XV), unsaturated azlactones (XVI) are formed, presumably by condensation of the aldehyde with the 2-phenyl-5-oxo-2-oxazoline formed first. Such unsaturated lactones (XVI) by standard conver-



sions furnish several kinds of useful products.

[W.J.CE.]

Bibliography: E D. Bergmann, The oxazolidines, *Chem Rev.*, 53:309–352, 1953; H. T. Clarke, J. R. Johnson, and R. Robinson (eds.), *The Chemistry of Penicillin*, 1949; R. C. Elderfield (ed.), *Heterocyclic Compounds*, vol. 5, 1957

Oxidation process

Literally, oxidation is the reaction of an element or compound with oxygen However, in the broadest sense, it is an increase in the valence of an atom or ion as a result of the loss of electrons. *See* OXIDATION-REDUCTION.

Many common oxidation processes lead to stable oxides such as metal oxides, carbon dioxide, water, and the oxides of sulfur, nitrogen, and phosphorus. They are typified in the natural processes of corrosion, decay, respiration, and combustion Controlled by man, they are part of the foundation of the heavy chemicals industry. The most valuable to the organic chemical industry are the partial oxidation products obtainable by careful control of the process. From the chemical engineering viewpoint, the oxidant is the most significant variable by which oxidation is controlled. The oxygen of the air is the ultimate oxidizing agent in most oxidations. It is used directly wherever sufficient control of the other variables yields the desired product economically. Where this is not possible, its action may be modified by conversion to some other form, such as to an oxidizing agent which shows greater selectivity in its attack on the other molecules.

Frequently, spent oxidant can be regenerated by contact with air or oxygen under reaction conditions similar to those under which it is employed. In cases where minute amounts of an oxidant can be utilized with air or oxygen, the oxidant is said to be a catalyst. Catalysis continues to be one of the most lucrative methods of controlling oxidation reactions to preselected products With proper catalyst selection, elemental oxygen may then be substituted for more expensive oxidants to give simpler processes.

Oxidations often are induced by means other than true catalysts, such as by energy or by high velocity particles. The entire spectrum of radiant energy, including heat, light, α-, β-, and γ radiation, has found applications in the field of oxidation. The extreme example, in which energy becomes the prime source of the oxidant, is electrolytic oxidation. The oxidant is generated in situ by application of electrical energy. *See* ELECTROCHEMICAL PROCESS.

Bacteria, aided by enzymes acting as catalysts, may consume oxygen and transform material to a higher state of oxidation. Fermentation, sewage sludge digestion, and acetic acid production from ethanol utilize such agents.

[illegible] represent the most

this category. Functionally, air and oxygen are interchangeable. Selection depends upon cost, heat removal, reaction rate, and product recovery problems. Classification is made into inorganic and organic oxidations. On the basis of the physical state at reaction temperature, oxidations are effected in the vapor, liquid, or solid phase.

Inorganic processes. These are typified by their high heat evolution. They are essentially combustion reactions, but because most are carried to the highest stable oxidation state, process control is simplified

Vapor-phase oxidations are used to produce the major heavy chemicals, for example, air oxidation of hydrogen sulfide or sulfur dioxide to sulfuric acid, of ammonia to nitric acid, of phosphorus vapor to phosphoric acid, of hydrogen chloride to chlorine, and of vaporized zinc to zinc oxide.

Liquid-phase oxidations of inorganic compounds are rare because so few are liquids. Liquid sulfur is burned to sulfur dioxide. At high temperatures, mercuric oxide is made from its elements, and litharge from molten lead Air and oxygen, blown through molten iron, make steel by oxidizing such impurities as carbon, sulfur, and phosphorus

Solid-phase oxidations are applied most commonly to obtain oxides from the elements. High purity carbon dioxide is made from coke in this way Oxidation of magnesium and aluminum have been considered in rocket and jet fuels. Lower oxides are converted to higher ones. Mixed lead oxides are purified to the monoxide litharge by roasting. Barium peroxide forms from the oxide. Two of the more powerful and costly inorganic oxidizing agents are obtained by processes involving gas-solid phase reactions Potassium permanganate is produced by roasting a mixture of manganese dioxide and potassium hydroxide with air in a kiln or muffle furnace. In an analogous way, chromite ore and sodium carbonate yield sodium chromate.

Organic processes. This group includes the oxidation of organic raw [illegible] ials to their potential oxidation product [illegible] nd no [illegible]

both in vapor and liquid phase, are em- ...termediate products, all of which are in commercial quantities, include olefins, , alcohols, hydroperoxides, epoxides, , ketones, acids, anhydrides, and esters. ...hase processes. Uncontrolled, most or- ...npounds will oxidize by combustion to value products as CO, CO_2, and water. ...of interest only where energy is the de- ...luct. Other products may be obtained by ...alysts and by exercising control over the ...ables

...noncatalytic process, control of the prod- ...rovided by diluting the reaction mixture. ...ydrocarbons such as propane or butane, ...ith air and diluents of recycling spent ...e permitted to react in empty tubes under ...00 psi and 700–900°F, and are then ...l with an aqueous product stream such as ...hyde The products in solution consist of ...different products of value An elaborate ... system isolates methanol, formaldehyde, ...hyde, acetone, aliphatic acids, and mixed ... Others present include n-propyl and butyl ..., methyl ethyl ketone, and oxides of lower

...ame principle is used to produce acetylene ...ethane by the S...

...ed with water after extremely short reaction (less than 0.01 sec) to catch the relatively ...le acetylene before it decomposes further ...different reaction conditions, methane yields ... By burning in a limited air supply and ...ing the products by impingement on chan- ...ns, channel black is formed

...tures, such as those from noncatalyzed ... oxidation, are avoided by oxidizing com- ...s selected for their reactivity toward oxygen ...presence of catalysts Thus, ethylene, when ...with oxygen and passed over a silver catalyst, ... ethylene oxide by direct addition Many ...ols, when treated similarly over silver or cop- ...are dehydrogenated to aldehyde or ketones. ...oxygen in this case combines only with the ...sed hydrogen to form water In this way, ...aldehyde is made from methanol, acetaldehyde ... ethanol, and acetone from isopropyl alcohol ... addition of oxygen and dehydrogenation can ...nduced to occur at activated positions in a ...cule by the proper catalysts Propylene, which ...it be expected to be attacked at the double ...l is oxidized to acrolein over copper- and ...mium-containing catalysts, because of the acti- ...ng effect of the double bond on the allylic car- ...hydrogen bond The benzene ring has a similar ...ct on the methyl groups of toluene, so that ...aldehyde can be produced in commercial ...ntities The aromatic ring in this case not only ...ates the side chain, but is itself stable toward ...dation. This stability, combined with the selec- ... catalytic effect of the oxides of vanadium, has led to the commercial processes which oxidize naphthalene and o-xylene to phthalic anhydride. The volatility and stability of this cyclic anhydride helps prevent its further oxidation before issuing from the reactor. At higher temperatures, even the benzene ring will disintegrate over this catalyst, giving commercially attractive yields of maleic anhydride from benzene itself

The air oxidation of ammonia to nitric oxides, themselves powerful oxidants, can be combined with the oxidation of organic compounds to yield the nitrile grouping. Hydrogen cyanide is now produced on a large scale from methane, ammonia, ...

limits the amount of oxygen which contacts the organic system at any one time Increased control of heat of reaction and conversions is thus possible. Catalysts are usually used in solution to avoid problems of coking and regeneration. Solvents can be employed, often with strong influences on the oxidations

Most noncatalytic oxidations are autocatalytic; once started they may be self-perpetuating They are chain reactions in which free radicals are formed and consumed continually in a series of steps involving both oxygen and hydrocarbons. The initiating step is open to question, but may involve the small amount of ozone occurring in air or oxygen Hydroperoxides are the first *stable species isolable* In the early 1950s, a commercial process was developed to produce phenol from cumene by first forming the hydroperoxide, which is then cleaved with acid to the phenol and acetone. The same grouping forms in acetaldehyde oxidation, giving peracetic acid, an extremely useful reagent Hydrogen peroxide itself is now made commercially by the autoxidation of 2-ethyl hydroanthraquinone in an organic solvent from which the peroxide is extracted The resultant ethyl anthraquinone is reduced with hydrog...

F...

fun...

to new species, as in the case of cumene hydroperoxide to phenol Metals and their oxides and salts can be used to favor aldehyde or ketone formation Because the former are easily oxidized to acids, in one operation a variety of hydrocarbons can be converted catalytically to carboxylic acids. Benzoic and t-butylbenzoic acids are obtained from toluene and t-butyltoluene, respectively The toluic and phthalic acids are formed from their respective xylenes Liquid-phase butane oxidation again gives mixed products, including appreciable quantities of acetic acid Paraffin wax oxidizes to fatty acids, which were used in World War II to make synthetic soap and butter in Germany. Secondary hydroperoxides decompose to ketones instead of aldehydes under such conditions These are more resistant to oxidation Acetophenone is produced from ethylbenzene and cyclohexanone from cyclohexane More vigorous oxidation will cleave car-

bon-carbon bonds to give benzoic acid from the former and adipic acid from the latter.

Chemical oxidants. These can be highly selective, although more than one may be capable of effecting a given oxidation. The choice depends upon the cost, scale of operations, availability, groups oxidized, product desired, by-products, ease of recovery, corrosion, and product purification. The most widely utilized are summarized, with emphasis on commercially practical processes.

Nitric acid. In concentrations of about 30% or lower, the nitration capacity of nitric acid becomes minor compared to its oxidizing power It is used to produce carboxylic acids. Cyclohexanol is cleaved to adipic acid. Paraxylene and *p*-toluic acid yield terephthalic acid. Both are important building blocks of high polymers.

Peroxides Hydrogen peroxide is used alone or in mixtures where peracids exist in equilibrium with it For instance, hydrogen peroxide in acetic acid is in equilibrium with peracetic acid which is most effective in the epoxidation and hydroxylation of double bonds Glycols are formed from monoolefins; glycerine from allyl alcohol *See* PEROXIDE

Ozone. Alone or with oxygen, ozone is most active toward unsaturation. Highly unstable ozonides may be produced, but under controlled conditions, this agent can be highly selective to produce alcohols, aldehydes, and esters, usually with carbon-carbon scission Castor oil is cleaved to a dibasic and a monobasic acid via ozonization. *See* OZONIZATION

Sulfates. Fuming sulfuric acid will attack even saturated hydrocarbons at high temperatures. Controlled properly, as in the presence of mercuric salts, it was once used to produce phthalic anhydride from naphthalene. Sulfur dioxide is evolved In the middle 1950s, neutral sulfates such as ammonium sulfate were found to be commercially suitable as oxidants to manufacture iso- and terephthalic acids from the xylenes

Permanganate. Potassium permanganate has been used widely for a long time where cost was not important and where an activated position existed in the compound to be oxidized Olefins yield glycols; methylbenzenes carboxylic acids.

Of less economic significance in chemical processing are such agents as hypochlorite (bleaching), dichromates (pickling), chlorates, sulfur and its oxides (Willgerodt reaction), copper salts (phenols from aromatics), lead tetraacetate (sugar chemistry), oxides of silver, selenium, manganese, mercury, and iron, and high-temperature aqueous alkali. *See* BLEACHING; CATALYSIS; EPOXIDATION, UNIT PROCESSES. [W.G.T.]

Bibliography: T. Brooks (ed.), *Chemistry of Petroleum Hydrocarbons*, vol 2, 1955; P. H. Groggins (ed.), *Unit Processes in Organic Synthesis*, 5th ed , 1958.

Oxidation-reduction

An important concept of chemical reactions which is useful in systematizing the chemistry of many substances. Oxidation can be represented as involving a loss of electrons by one substance, and reduction as involving an absorption of electrons by another substance. Both oxidation and reduction occur simultaneously and in equivalent amounts during any reaction involving either process.

Some important reactions of common occurrence which involve oxidation are: rusting of iron or corrosion of metals in general, combustion of hydrocarbons, the reaction of carbohydrates with oxygen (this takes place in a controlled manner in living cells). The following are some reactions which involve reduction: transformation of carbon dioxide to carbohydrates (this takes place in photosynthesis), the winning of metals from oxides, sulfides, or halides, electrodeposition of metals, hydrogenation of fats

The oxidation number. The oxidation number, or oxidation state, is a concept used to describe some important aspects of the state of combination of the elements. It specifies whether the element in question is combined with elements which are more electropositive or more electronegative than it is. It further specifies the combining capacity which the element exhibits in a particular combination. A scale of oxidation numbers is defined by assigning to the oxygen atom in its usual state of combination with other atoms, the value of −2. Applying this definition, and the added one that the sum of the oxidation numbers for the atoms comprising a particular unit equals the net charge on that unit, the oxidation number of iron. Fe, in Fe_2O_3 and in Fe^{+++} is seen to be +3, and of sulfur, S, in SO_3 and in SO_4^{--} is seen to be +6. The oxidation numbers of the positive centers in turn can be used to establish those of other electronegative atoms with which they are combined

The scale of oxidation numbers which has been defined has not been chosen arbitrarily. It recognizes that oxygen is electronegative toward other elements (the single exception being fluorine); it also recognizes that, when oxygen reacts with other elements, each atom seeks to acquire two additional electrons, by sharing or transfer, to complete a stable valence shell of 8 electrons A tidier definition of the scale would be possible if fluorine were made the standard and assigned the value of −1 in the combined state The advantage this would offer is that fluorine is the most electronegative of all atoms and exhibits only a single combining capacity, so that fewer exceptions in definition would be necessary than are required for oxygen For example, to be consistent with the general philosophy of the concept of oxidation number, oxygen in fluorine oxide, F_2O, must be assigned the value +2, and in peroxide the value of −1 (In hydrogen peroxide, H_2O_2, the oxygen has not exhausted its full combining capacity for hydrogen, and the reaction $H_2O_2 + H_2 = 2H_2O$ is possible) Even with these exceptions, oxygen is more serviceable as a standard because of the large number of substances of which it is a part.

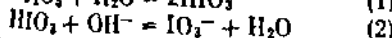
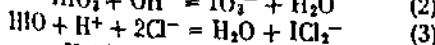
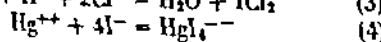
The oxidation number by no means gives a complete description of the state of combination of an atom. Thus it makes no [illegible] n between fluorine

in HF, AlF_3, and NaF, although the actual electric charges on the fluorine atom in these compounds will differ appreciably. The utility of the concept is based in part on this feature because much of the chemistry of these substances can be understood when it is realized that each of them readily yields F^-, as is the case when they dissolve in water. The chemistry of the three substances, regarding the fluorine part of the combinations, is concerned with reactions of F^-. Although oxidation number is in some respects similar to valence, the two concepts have distinct meanings. In the substance H_2, the valence of hydrogen is 1 because each H makes a single bond to another H, but the oxidation number is 0, because the hydrogen is uncombined with a different element. *See* VALENCE.

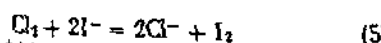
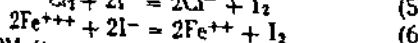
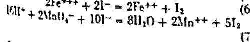
The systematization of chemistry based on the concept of oxidation number can be illustrated with reference to the chemistry of iodine. The usual oxidation states exhibited by iodine are −1, 0, +1, +5, and +7. Examples of substances corresponding to each oxidation state are

+7	IO_4^-, HIO_4, IF_7
+5	I_2O_5, IO_3^-, HIO_3, IF_5
+1	HIO, IO^-, ICl_2^-
0	I_2
−1	I^-, HI, NaI

Following the classification by oxidation number the substances *containing* the element in question, the reactions of the substances fall naturally into two classes. In the first class, no change in oxidation number takes place, and in the second, the class of oxidation-reduction reactions, changes in oxidation number do take place. Some examples of the first class are


$$I_2O_5 + H_2O = 2HIO_3 \quad (1)$$

$$HIO_3 + OH^- = IO_3^- + H_2O \quad (2)$$

$$HIO + H^+ + 2Cl^- = H_2O + ICl_2^- \quad (3)$$

$$Hg^{++} + 4I^- = HgI_4^{--} \quad (4)$$

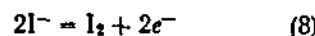
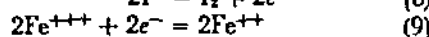
and of the second are


$$Cl_2 + 2I^- = 2Cl^- + I_2 \quad (5)$$

$$2Fe^{+++} + 2I^- = 2Fe^{++} + I_2 \quad (6)$$

$$16H^+ + 2MnO_4^- + 10I^- = 8H_2O + 2Mn^{++} + 5I_2 \quad (7)$$

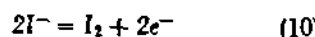
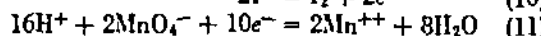
In reactions of the first class, some center regarded as positive undergoes a change in the nature of the groups associated with it, but without undergoing a change in oxidation number. Reaction (3), for example, describes th[illegible] by 2 Cl [illegible] in oxidat[illegible] be accomp[illegible] by changes in the state of association of the centers in question.

Reactions (5), (6), and (7) illustrate the utility of the concept of oxidation number. A variety of reagents as different in their properties as Cl_2, Fe^{+++}, and MnO_4^- serve to bring about the change from I^- to I_2. However, their chemical individuality does not affect the state of the iodine, and no group characteristic of the oxidizing agent is necessarily transferred in the net change. This situation obtains only for reactions in a strongly solvating medium such as water, which provides the groups that associate with the species being considered. Thus, when the reactions take place in the solid, it is necessary to specify what iodide is being used, whether sodium iodide, NaI, or silver iodide, AgI, for example, and the products would be strongly dependent on the choice.

Oxidation-reduction reactions. In an oxidation-reduction reaction, some element decreases in oxidation state and some element increases in oxidation state. The substances containing these elements are defined as the oxidizing agents and reducing agents, and they are said to be reduced and oxidized, respectively. The processes in question can always be represented formally as involving electron absorption by the oxidizing agent and electron loss by the reducing agent. For example, reaction (6) can be regarded as the sum of the two partial processes or half-reactions:


$$2I^- = I_2 + 2e^- \quad (8)$$

$$2Fe^{+++} + 2e^- = 2Fe^{++} \quad (9)$$

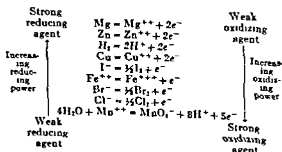
Similarly, reaction (7) consists of the two half-reactions:


$$2I^- = I_2 + 2e^- \quad (10)$$

$$16H^+ + 2MnO_4^- + 10e^- = 2Mn^{++} + 8H_2O \quad (11)$$

with half-reaction (10) being taken five times to balance the electron flow from reducing agent to oxidizing agent.

Each half reaction consists of an oxidation-reduction couple; thus, in half reaction (11), the reducing agent and oxidizing agent comprising the couple are manganous ion, Mn^{++}, and permanganate ion, MnO_4^-, respectively; in (10), the reducing agent is I^-, and the oxidizing agent, I_2. The direction of reaction (7) implies that MnO_4^- in acid solution is a stronger oxidizing agent than is I_2. Because of the reciprocal relation between the oxidizing agent and reducing agent compris[illegible] ple [illegible] a s [illegible] tion. [illegible] agents may be ranked in order of tendency to react, and this ranking immediately implies an opposite order of tendency to react for the [illegible]

[illegible]



A complete list contains the displacement series of the metals. The most powerful reducing agent

bon-carbon bonds to give benzoic acid from the former and adipic acid from the latter.

Chemical oxidants. These can be highly selective, although more than one may be capable of effecting a given oxidation The choice depends upon the cost, scale of operations, availability, groups oxidized, product desired, by-products, ease of recovery, corrosion, and product purification. The most widely utilized are summarized, with emphasis on commercially practical processes.

Nitric acid. In concentrations of about 30% or lower, the nitration capacity of nitric acid becomes minor compared to its oxidizing power. It is used to produce carboxylic acids. Cyclohexanol is cleaved to adipic acid. Paraxylene and *p*-toluic acid yield terephthalic acid. Both are important building blocks of high polymers.

Peroxides. Hydrogen peroxide is used alone or in mixtures where peracids exist in equilibrium with it. For instance, hydrogen peroxide in acetic acid is in equilibrium with peracetic acid which is most effective in the epoxidation and hydroxylation of double bonds. Glycols are formed from monoolefins; glycerine from allyl alcohol. *See* PEROXIDE.

Ozone. Alone or with oxygen, ozone is most active toward unsaturation. Highly unstable ozonides may be produced, but under controlled conditions, this agent can be highly selective to produce alcohols, aldehydes, and esters, usually with carbon-carbon scission Castor oil is cleaved to a dibasic and a monobasic acid via ozonization. *See* OZONIZATION.

Sulfates Fuming sulfuric acid will attack even saturated hydrocarbons at high temperatures. Controlled properly, as in the presence of mercuric salts, it was once used to produce phthalic anhydride from naphthalene. Sulfur dioxide is evolved. In the middle 1950s, neutral sulfates such as ammonium sulfate were found to be commercially suitable as oxidants to manufacture iso- and terephthalic acids from the xylenes

Permanganate. Potassium permanganate has been used widely for a long time where cost was not important and where an activated position existed in the compound to be oxidized. Olefins yield glycols; methylbenzenes carboxylic acids

Of less economic significance in chemical processing are such agents as hypochlorite (bleaching), dichromates (pickling), chlorates, sulfur and its oxides (Willgerodt reaction), copper salts (phenols from aromatics), lead tetraacetate (sugar chemistry), oxides of silver, selenium, manganese, mercury, and iron, and high-temperature aqueous alkali. *See* BLEACHING; CATALYSIS; EPOXIDATION; UNIT PROCESSES [W G.T]

Bibliography: T. Brooks (ed.), *Chemistry of Petroleum Hydrocarbons*, vol. 2, 1955; P H. Groggins (ed.), *Unit Processes in Organic Synthesis*, 5th ed., 1958.

Oxidation-reduction

An important concept of chemical reactions which is useful in systematizing the chemistry of many substances Oxidation can be represented as involving a loss of electrons by one substance, and reduction as involving an absorption of electrons by another substance. Both oxidation and reduction occur simultaneously and in equivalent amounts during any reaction involving either process

Some important reactions of common occurrence which involve oxidation are: rusting of iron or corrosion of metals in general, combustion of hydrocarbons, the reaction of carbohydrates with oxygen (this takes place in a controlled manner in living cells). The following are some reactions which involve reduction: transformation of carbon dioxide to carbohydrates (this takes place in photosynthesis), the winning of metals from oxides, sulfides, or halides, electrodeposition of metals, hydrogenation of fats.

The oxidation number. The oxidation number, or oxidation state, is a concept used to describe some important aspects of the state of combination of the elements. It specifies whether the element in question is combined with elements which are more electropositive or more electronegative than it is It further specifies the combining capacity which the element exhibits in a particular combination A scale of oxidation numbers is defined by assigning to the oxygen atom in its usual state of combination with other atoms, the value of −2. Applying this definition, and the added one that the sum of the oxidation numbers for the atoms comprising a particular unit equals the net charge on that unit, the oxidation number of iron, Fe, in Fe_2O_3 and in Fe^{+++} is seen to be +3, and of sulfur, S, in SO_3 and in SO_4^{--} is seen to be +6. The oxidation numbers of the positive centers in turn can be used to establish those of other electronegative atoms with which they are combined

The scale of oxidation numbers which has been defined has not been chosen arbitrarily. It recognizes that oxygen is electronegative toward other elements (the single exception being fluorine); it also recognizes that, when oxygen reacts with other elements, each atom seeks to acquire two additional electrons, by sharing or transfer, to complete a stable valence shell of 8 electrons A tidier definition of the scale would be possible if fluorine were made the standard and assigned the value of −1 in the combined state. The advantage this would offer is that fluorine is the most electronegative of all atoms and exhibits only a single combining capacity, so that fewer exceptions in definition would be necessary than are required for oxygen. For example, to be consistent with the general philosophy of the concept of oxidation number, oxygen in fluorine oxide, F_2O, must be assigned the value +2, and in peroxide the value of −1. (In hydrogen peroxide, H_2O_2, the oxygen has not exhausted its full combining capacity for hydrogen, and the reaction $H_2O_2 + H_2 = 2H_2O$ is possible.) Even with these exceptions, oxygen is more serviceable as a standard because of the large number of substances of which it is a part.

The oxidation number by no means gives a complete description of the state of combination of an atom. Thus it makes no distinction between fluorine

the chemistry of these substances can be understood when it is realized that each of them readily yields F^-, as is the case when they dissolve in water. The chemistry of the three substances, regarding the fluorine part of the combinations, is concerned with reactions of F^-. Although oxidation number is in some respects similar to valence, the two concepts have distinct meanings. In the substance H_2, the valence of hydrogen is 1 because each H makes a single bond to another H, but the oxidation number is 0, because the hydrogen is uncombined with a different element. See VALENCE.

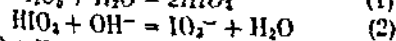
The systematization of chemistry based on the concept of oxidation number can be illustrated with reference to the chemistry of iodine. The usual oxidation states exhibited by iodine are −1, 0, +1, +5, and +7. Examples of substances corresponding to each oxidation state are

+7	IO_4^-, HIO_4, IF_7
+5	I_2O_5, IO_3^-, HIO_3, IF_5
+1	HIO, IO^-, ICl_2^-
0	I_2
−1	I^-, HI, NaI

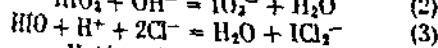
Following the classification by oxidation number of the substances containing the element in question, the reactions of the substances fall naturally into two classes. In the first class, no change in oxidation number takes place, and in the second, the class of oxidation-reduction reactions, changes in oxidation number do take place. Some examples of the first class are



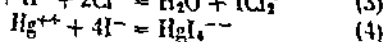
$$I_2O_5 + H_2O = 2HIO_3 \quad (1)$$



$$HIO_3 + OH^- = IO_3^- + H_2O \quad (2)$$

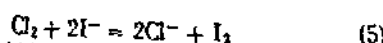


$$HIO + H^+ + 2Cl^- = H_2O + ICl_2^- \quad (3)$$

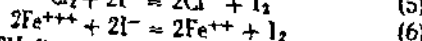


$$Hg^{++} + 4I^- = HgI_4^{--} \quad (4)$$

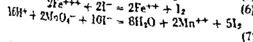
and of the second are



$$Cl_2 + 2I^- = 2Cl^- + I_2 \quad (5)$$



$$2Fe^{+++} + 2I^- = 2Fe^{++} + I_2 \quad (6)$$



$$16H^+ + 2MnO_4^- + 10I^- = 8H_2O + 2Mn^{++} + 5I_2 \quad (7)$$

In reactions of the first class, some center regarded as positive undergoes a change in the nature of the groups associated with it, but without undergoing a change in ...

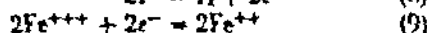
... ompanied by changes in the state of association of the centers in question.

Reactions (5), (6), and (7) illustrate the utility of the concept of oxidation number. A variety of reagents as different in their properties as Cl_2, Fe^{+++}, and MnO_4^- serve to bring about the change from I^- to I_2. However their chem... does not affe... characterist... transferred ... this situation obtains only for reactions in a strongly solvating medium such as water, which provides the groups that associate with the species being considered. Thus, when the reactions take place in the solid, it is necessary to specify what iodide is being used, whether sodium iodide, NaI, or silver iodide, AgI, for example, and the products would be strongly dependent on the choice.

Oxidation-reduction reactions. In an oxidation-reduction reaction, some element decreases in oxidation state and some element increases in oxidation state. The substances containing these elements are defined as the oxidizing agents and reducing agents, and they are said to be reduced and oxidized, respectively. The processes in question can always be represented formally as involving electron absorption by the oxidizing agent and electron loss by the reducing agent. For example, reaction (6) can be regarded as the sum of the two partial processes or half reactions:



$$2I^- = I_2 + 2e^- \quad (8)$$

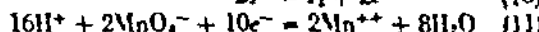


$$2Fe^{+++} + 2e^- = 2Fe^{++} \quad (9)$$

Similarly, reaction (7) consists of the two half-reactions:



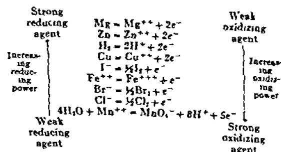
$$2I^- = I_2 + 2e^- \quad (10)$$



$$16H^+ + 2MnO_4^- + 10e^- = 2Mn^{++} + 8H_2O \quad (11)$$

with half-reaction (10) being taken five times to balance the electron flow from reducing agent to oxidizing agent.

Each half-reaction consists of an oxidation-reduction couple; thus, in half-reaction (11), the reducing agent and oxidizing agent comprising the couple are manganous ion, Mn^{++}, and permanganate ion, MnO_4^-, respectively; in (10), the reducing agent is I^-, and the oxidizing agent, I_2. The direction of reaction (7) implies that MnO_4^- in acid solution is a stronger oxidizing agent than is I_2. Because of the reciprocal relation between the oxidizing agent and reducing agent comprising a couple, this statement is equivalent to saying that I^- is a stronger reducing agent than Mn^{++} in acid solution. Reducing agents may be ranked in order of tendency to react, and this ranking immediately implies an opposite order of tendency to react for the oxidizing agents which complete the couples. A list containing some oxidation-reduction couples ordinarily encountered and ranked in such fashion follows:



A complete list contains the displacement series of the metals. The most powerful reducing agent

bon-carbon bonds to give benzoic acid from the former and adipic acid from the latter.

Chemical oxidants. These can be highly selective, although more than one may be capable of effecting a given oxidation The choice depends upon the cost, scale of operations, availability, groups oxidized, product desired, by-products, ease of recovery, corrosion, and product purification. The most widely utilized are summarized, with emphasis on commercially practical processes.

Nitric acid. In concentrations of about 30% or lower, the nitration capacity of nitric acid becomes minor compared to its oxidizing power. It is used to produce carboxylic acids. Cyclohexanol is cleaved to adipic acid Paraxylene and *p*-toluic acid yield terephthalic acid. Both are important building blocks of high polymers.

Peroxides. Hydrogen peroxide is used alone or in mixtures where peracids exist in equilibrium with it. For instance, hydrogen peroxide in acetic acid is in equilibrium with peracetic acid which is most effective in the epoxidation and hydroxylation of double bonds. Glycols are formed from monoolefins; glycerine from allyl alcohol. *See* PEROXIDE.

Ozone. Alone or with oxygen, ozone is most active toward unsaturation Highly unstable ozonides may be produced, but under controlled conditions, this agent can be highly selective to produce alcohols, aldehydes, and esters, usually with carbon-carbon scission Castor oil is cleaved to a dibasic and a monobasic acid via ozonization. *See* OZONIZATION.

Sulfates. Fuming sulfuric acid will attack even saturated hydrocarbons at high temperatures Controlled properly, as in the presence of mercuric salts, it was once used to produce phthalic anhydride from naphthalene Sulfur dioxide is evolved In the middle 1950s, neutral sulfates such as ammonium sulfate were found to be commercially suitable as oxidants to manufacture iso- and terephthalic acids from the xylenes.

Permanganate. Potassium permanganate has been used widely for a long time where cost was not important and where an activated position existed in the compound to be oxidized. Olefins yield glycols; methylbenzenes carboxylic acids.

Of less economic significance in chemical processing are such agents as hypochlorite (bleaching), dichromates (pickling), chlorates, sulfur and its oxides (Willgerodt reaction), copper salts (phenols from aromatics), lead tetraacetate (sugar chemistry), oxides of silver, selenium, manganese, mercury, and iron, and high-temperature aqueous alkali. *See* BLEACHING; CATALYSIS; EPOXIDATION; UNIT PROCESSES. [W.C.T.]

Bibliography: T. Brooks (ed.), *Chemistry of Petroleum Hydrocarbons*, vol. 2, 1955; P. H. Groggins (ed.), *Unit Processes in Organic Synthesis*, 5th ed., 1958.

Oxidation-reduction

An important concept of chemical reactions which is useful in systematizing the chemistry of many substances. Oxidation can be represented as involving a loss of electrons by one substance, and reduction as involving an absorption of electrons by another substance. Both oxidation and reduction occur simultaneously and in equivalent amounts during any reaction involving either process.

Some important reactions of common occurrence which involve oxidation are: rusting of iron or corrosion of metals in general, combustion of hydrocarbons, the reaction of carbohydrates with oxygen (this takes place in a controlled manner in living cells). The following are some reactions which involve reduction: transformation of carbon dioxide to carbohydrates (this takes place in photosynthesis), the winning of metals from oxides, sulfides, or halides, electrodeposition of metals, hydrogenation of fats.

The oxidation number. The oxidation number, or oxidation state, is a concept used to describe some important aspects of the state of combination of the elements. It specifies whether the element in question is combined with elements which are more electropositive or more electronegative than it is It further specifies the combining capacity which the element exhibits in a particular combination A scale of oxidation numbers is defined by assigning to the oxygen atom in its usual state of combination with other atoms, the value of −2. Applying this definition, and the added one that the sum of the oxidation numbers for the atoms comprising a particular unit equals the net charge on that unit, the oxidation number of iron, Fe, in Fe_2O_3 and in Fe^{+++} is seen to be +3, and of sulfur, S, in SO_3 and in SO_4^{--} is seen to be +6. The oxidation numbers of the positive centers in turn can be used to establish those of other electronegative atoms with which they are combined

The scale of oxidation numbers which has been defined has not been chosen arbitrarily. It recognizes that oxygen is electronegative toward other elements (the single exception being fluorine); it also recognizes that, when oxygen reacts with other elements, each atom seeks to acquire two additional electrons, by sharing or transfer, to complete a stable valence shell of 8 electrons A tidier definition of the scale would be possible if fluorine were made the standard and assigned the value of −1 in the combined state. The advantage this would offer is that fluorine is the most electronegative of all atoms and exhibits only a single combining capacity, so that fewer exceptions in definition would be necessary than are required for oxygen For example, to be consistent with the general philosophy of the concept of oxidation number, oxygen in fluorine oxide, F_2O, must be assigned the value +2, and in peroxide the value of −1 (In hydrogen peroxide, H_2O_2, the oxygen has not exhausted its full combining capacity for hydrogen, and the reaction $H_2O_2 + H_2 = 2H_2O$ is possible) Even with these exceptions, oxygen is more serviceable as a standard because of the large number of substances of which it is a part.

The oxidation number by no means gives a complete description of the state of combination of an atom. Thus it makes n n betw

atter of quicklime (CaO) by the reaction



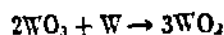
$$CaCO_3 \rightarrow CaO + CO_2$$

in which carbon dioxide is driven off Gold oxide may be prepared by heating gold hydroxide



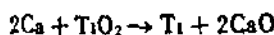
$$2Au(OH)_3 \rightarrow Au_2O_3 + 3H_2O$$

Higher oxides of an element may be reduced to lower oxides, usually at high temperatures, for example, the reduction of tungsten trioxide by tungsten



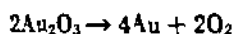
$$2WO_3 + W \rightarrow 3WO_2$$

Complete reduction to the element may be performed by other elements whose oxides are more stable, as in the formation of calcium oxide from titanium dioxide by the reaction



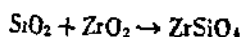
$$2Ca + TiO_2 \rightarrow Ti + 2CaO$$

Although the solid oxides of a few metals such as mercury and the noble metals can be easily decomposed by heating, for example,



$$2Au_2O_3 \rightarrow 4Au + 3O_2$$

most metal oxides are very stable and many constitute important refractory materials. For example, magnesium oxide, calcium oxide, and zirconium dioxide do not melt or vaporize appreciably at temperatures up to 2500°C. A great number of refractories consist of compounds of two or more oxides, silicon dioxide and zirconium dioxide form zirconium silicate by the reaction

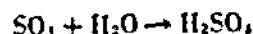


$$SiO_2 + ZrO_2 \rightarrow ZrSiO_4$$

Because so many oxides can be easily formed, studies of them have been most important in establishing relative atomic weights of the elements based on the defined atomic weight for oxygen. Furthermore, these studies were fundamental in forming the basis for the laws of definite proportions and multiple proportions for compounds It is of special significance that, although any gaseous oxide species must necessarily have a definite oxygen-to-element proportion, a number of solid and liquid oxides can be prepared with proportions which may vary continuously over a considerable range This is particularly true for oxides prepared under equilibrium conditions at high temperatures Thus, titanium exposed to oxygen until reaction equilibrium is reached at a number of selected conditions of temperature and oxygen pressure will form the solid oxide, TiO It has the same crystal structure as rock salt; that is, every other site along the three coordinate directions of the crystal will be occupied by titanium atoms and the alternate sites by oxygen atoms (each in its ion form Ti^{++} and O^{--}) to give the simple Ti/O ratio of 1·1 However, with other selected pressure temperature conditions, oxides of this same structure at every Ti/O ratio from 1.0.7 to 1·1 25 may be prepared The variable proportions are a manifestation that variable numbers of oxygen or titanium sites can simply remain vacant in a homogeneous way The range is referred to as the TiO/O 1:07–1.25 phase, or more loosely, the TiO solid solution phase.

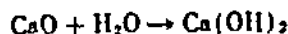
Most of the nonmetal oxides commonly encountered as gases, such as SO_2 and CO_2, form solids and liquids in which the molecular units of the gas are retained so that the simple definite proportions are clearly maintained. Such oxides melt and boil at low temperatures, because the molecular units are weakly bonded to adjoining molecular units.

Oxides may be classified as acidic or basic according to the character of the solution resulting from their reactions with water. The nonmetal oxides generally form acid solutions, for example,



$$SO_3 + H_2O \rightarrow H_2SO_4$$

for the formation of sulfuric acid. The metal oxides generally form alkaline solutions, for example,



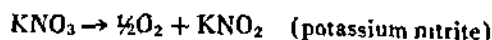
$$CaO + H_2O \rightarrow Ca(OH)_2$$

for the formation of calcium hydroxide or slaked lime However, given metals of the groups IV and higher of the periodic table will often have basic, intermediate, and acidic oxides Here, the acid character increases with increasing oxygen-metal ratio *See* ACID AND BASE; EQUIVALENT WEIGHT; OXYGEN; REFRACTORY. [R K E]

Bibliography: L. Brewer, The thermodynamic properties of the oxides and their vaporization processes, *Chem Rev.*, 52(1):1–74, 1953; M. Hansen, *Constitution of Binary Alloys*, 2d ed., 1958; P C. L. Thorne and E. R. Roberts (eds.), *Inorganic Chemistry*, 6th ed , 1955.

Oxidizing agent

A participant in a chemical reaction which accepts electrons from another reactant. In acting as an oxidizing agent, the substance undergoes a loss in positive valence or a gain in negative valence. In a more restricted sense, it is a species which supplies oxygen to another reactant. This is the more familiar concept of an oxidizing agent. Some of the more important oxidizing agents in this latter sense include hydrogen peroxide (H_2O_2), permanganate ion (MnO_4^-), potassium chlorate ($KClO_3$), dichromate ion ($Cr_2O_7^{--}$), nitric acid (HNO_3), hypochlorite ion (ClO^-), and potassium nitrate (KNO_3) An example of the action of such an oxidizing agent is that of potassium nitrate on carbon When potassium nitrate is heated in a porcelain crucible, oxygen is liberated:



$$KNO_3 \rightarrow \tfrac{1}{2}O_2 + KNO_2 \quad \text{(potassium nitrite)}$$

If finely divided carbon is present, it will be ignited and burn to carbon dioxide:



$$C + O_2 \rightarrow CO_2$$

An example of an oxidizing agent in the less restricted sense is that involving the reaction of chlorine (Cl_2) and ferrous chloride ($FeCl_2$):



$$FeCl_2 + \tfrac{1}{2}Cl_2 \rightarrow FeCl_3$$

Chlorine undergoes a change in valence from zero to 1–. It therefore acts as an oxidizing agent and is reduced in the process.

shown in the list is magnesium, Mg, although this is not the most powerful known. Magnesium is capable of reacting with any oxidizing agent below it in the list to yield Mg^{++} and to form the reduced product resulting from the oxidizing agent. Similarly, permanganate ion, MnO_4^-, in acid, the strongest oxidizing agent shown, is capable of reacting with any reducing agent above it in the list. Conversely, an oxidizing agent at the top of the list will not react appreciably with the reducing agent of a couple below it. The list given, containing 9 entries, can be used to predict the results of 70 separate experiments ($Mg + Zn^{++}$ on the one hand and $Mg^{++} + Zn$ on the other would be counted as separate experiments in arriving at this figure) *See* ELECTROCHEMICAL SERIES; ELECTRONEGATIVITY.

Since the driving force for a reaction depends on concentrations, the concentrations of all reactants and products must be specified, as well as other conditions, in applying a list such as that given. The order shown obtains for water solutions at 25°C, approximately 1 *M* in all soluble reagents and having gases present at approximately 1 atm pressure. A second limitation on the use of this list lies in the fact that it applies only when the expected reaction products are compatible Although copper is capable in principle of reducing iodine to form I^- and Cu^{++} at high concentration, these products are not compatible with each other, but they react to form copper(I) iodide, CuI Allowance for such features can always be made by incorporating the necessary additional half-reactions into the list. Finally, it must be stressed that the list can be used to predict the results of experiments only for systems which reach equilibrium sufficiently rapidly; it does not serve to predict the rate of reaction. To achieve the reduction of Fe^{+++} by H_2 predicted in the list, it would be necessary to use a catalyst in order to produce the reaction in a reasonable time.

The equilibrium information implied by a table of half-reactions can readily be put into quantitative form. Thus, the standard free energy change for the reaction of 1 equivalent weight of each reducing agent with some common oxidizing agent can be entered opposite each half-reaction. The numerical values of these quantities will be in the same order as are the half-reactions and can be combined algebraically to yield the standard free energy change, and therefore the equilibrium constant, for any reaction which can be written from the table. *See* EQUILIBRIUM, CHEMICAL.

A chemist concerned with reactions of the type under discussion will have a ready vocabulary of facts concerning oxidizing agents and reducing agents, such as their oxidizing or reducing powers, the speed with which they react, and the characteristics which may complicate their application. A typical problem in analytical chemistry is to reduce $Cr_2O_7^{--}$ to Cr^{+++} in acidic (perchloric acid) solution without introducing elements which are not already present. Metallic reducing agents such as zinc and iron or metal ion reducing agents are immediately eliminated from consideration because the products of oxidation may be difficult to remove from the resulting solution A solution of hydrogen iodide, HI, would be suitable, except that it would be necessary to take special pains to add it in equivalent amount because excess HI would be difficult to remove (the iodine, I_2, produced by oxidation of I^-, however, is easy to remove by extracting it with a suitable solvent such as carbon tetrachloride). Hydrogen would be ideal (the product of its oxidation, H^+, is already present in solution, and excess reducing agent can easily be removed) except that the rate of reaction would be disappointingly slow. A suitable reducing agent for this application would be hydrogen peroxide, H_2O_2; it reacts rapidly, the product of oxidation is oxygen, which is readily removed from solution, and excess oxidizing agent is easily destroyed by heating the solution. *See* ELECTRODE POTENTIAL; METABOLISM. [H T]

Bibliography: T. Moeller, *Inorganic Chemistry*, 1952.

Oxide

A binary compound of oxygen with another element. Oxides have been prepared for essentially all the elements except the noble gases Often, several different oxides of a given element can be prepared; a number exist naturally in the earth's crust and atmosphere: silicon dioxide (SiO_2) in quartz; aluminum oxide (Al_2O_3) in corundum; iron oxide (Fe_2O_3) in hematite; carbon dioxide (CO_2) gas, and water (H_2O)

Most elements will react with oxygen at appropriate temperature and oxygen pressure conditions, and many oxides may thus be directly prepared Phosphorus burns spontaneously in oxygen to form phosphorus pentoxide, $(P_2O_5)_2$. Sulfur requires ignition and thereafter burns to sulfur dioxide (SO_2) gas if the supply of oxygen is limited The relative amounts of oxygen and element available often determine which of several oxides will form, in an excess of oxygen, sulfur burns to form some sulfur trioxide (SO_3) gas Most metals in massive form react with oxygen only slowly at room temperatures because the first thin oxide coat formed protects the metal, magnesium and aluminum remain metallic in appearance for long periods because their oxide coatings are scarcely visible. However, diffusion of the oxygen and metal atoms through the film becomes rapid at high temperatures and these metals will burn intensely to their oxides if ignited. The oxides of the alkali and alkaline-earth metals, except for beryllium and magnesium, are porous when formed on the metal surface, and they provide only limited protection to the continuation of oxidation, even at room temperatures. Gold is exceptional in its resistance to oxygen, and its oxide (Au_2O_3) must be prepared by indirect means. The other noble metals, although ordinarily resistant to oxygen, will react at high temperatures to form gaseous oxides

Indirect preparation of the oxides may be accomplished by heating h s, nitrates, oxalates, or carbonates, a tion

... the ac-
reduced
ldoximes
See AL-
[L B C.]

HYDE, KETONE.

Oximetry

... technique for measurement of the fraction of ... hemoglobin in blood which is in oxygenated ...rm This fraction usually is expressed in per cent, and this percentage value is referred to as the oxygen saturation of blood The oximeter, the instrument used in this procedure, is a photoelectric photometer

One type of oximeter is designed to measure the ... portion ... e ... is

pinna of the ear, and the instrument used for this purpose is called an ear oximeter *See* PHOTOMETER.

A second type of oximeter is designed to measure the oxygen saturation of blood outside the body during or shortly after withdrawal of the blood from various sites in the vascular system Such a device usually is called a cuvette oximeter

The physical basis of oximetry stems from the difference in absorption by oxygenated and reduced hemoglobin of red light of wavelengths in the region of 640 mμ (Fig 1) These measurements of light absorption usually have been made on light transmitted through blood (transmission oximetry), but reflected light (reflection oximetry) also has been used successfully for this purpose

The first oximeters used successfully were relative reading devices, that is they were capable of measuring only changes in oxygen saturation and not absolute saturation The more recently devised

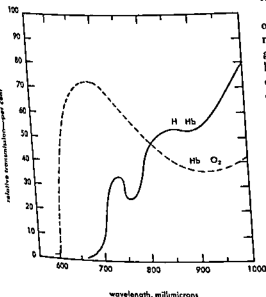


Fig 1 Comparison of spectral transmission of oxyhemoglobin (HbO_2) and reduced hemoglobin (HHb)

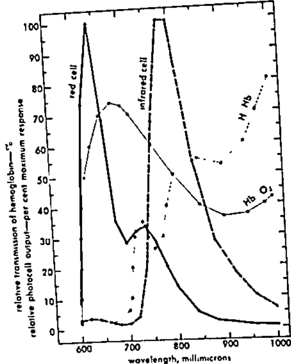


Fig 2 Comparison of spectral transmission of oxyhemoglobin and reduced hemoglobin and spectral sensitivities of oximeter photoelectric cells

absolute-reading oximeters of the transmission type determine the absorption by blood of light of two wavelengths, one of which (800 mμ) is absorbed equally by oxygenated and reduced hemoglobin, and the other of which (640 mμ) is absorbed very differently by these two forms of hemoglobin. These measurements are made by two photo-cell-filter assemblies each designed to have a peak sensitivity at one of these two wavelengths. (Fig 2)

Absolute measurement of the oxygen saturation of blood circulating in tissue, such as the ear, also requires some method for correction for the background spectral absorption of the tissues other than blood which necessarily must be interposed in the optical path of the instrument. This has been accomplished by incorporating a pneumatic pressure capsule in the earpiece which, when inflated to 200 mm Hg(mercury), presses the blood from the transilluminated portion of the ear so that the light absorption of the bloodless tissue can be determined in addition to the absorption of the normal, blood-containing tissue. The actual light absorption of the blood alone can then be determined by difference Such an oximeter earpiece is illustrated in Fig 3 A diagram of a cuvette oximeter is shown in Fig 4, along with the electrical circuity which frequently has been used with these devices In Fig 4*b*, control switch position 1 is the setting for reading the galvanometer zeros; control switch position 2 is used to adjust the sensitivity of the infrared cell for a single-scale operation when the earpiece is on the flushed ear or blood in the cuvette oximeter, control switch position

The terms, oxidizing agent and reducing agent, are relative; whether a substance acts as an oxidizing or reducing agent depends upon the nature and concentration of the reactant with which it is brought into contact.

The behavior in any oxidation-reduction reaction depends strongly upon the concentration. For certain reactant pairs, the roles of oxidizing agent and reducing agent may be interchanged by altering the concentration ratio. *See* ELECTROCHEMICAL SERIES; ELECTRONEGATIVITY; OXIDATION-REDUCTION; VALENCE. [F.J.J.]

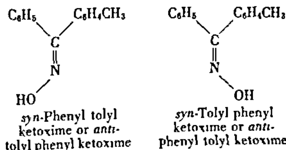
Oxime

One of a group of chemicals derived from aldehydes (RCH=NOH, aldoximes) or ketones (RR'C=NOH, ketoximes) used for isolation and identification of carbonyl compounds. In general, they are easily purified and have characteristic melting points. The properties of certain oximes have made them industrially important.

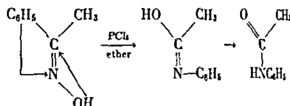
Oximes have received considerable attention because of their stereochemistry and their participation in the Beckmann rearrangement.

The discovery of geometrical isomers involving a carbon-nitrogen double bond demonstrated the fact of restricted rotation about such a bond in a manner analogous to that obtaining about a carbon-carbon double bond. *See* ISOMERISM, MOLECULAR. However, relatively few pairs of geometric isomers of the oximes, which are conventionally termed syn and anti isomers analogous to the more familiar cis and trans terminology used in carbon-carbon systems, have been isolated. This suggests that interconversion of such isomers involves relatively little energy. Thus, *syn*-benzaldehyde oxime (H and OH in a cis arrangement with respect to the double bond) is converted to the anti (trans) form by ethereal hydrogen chloride solution; reversion to the syn form can be accomplished by irradiation of a benzene solution of the anti form.

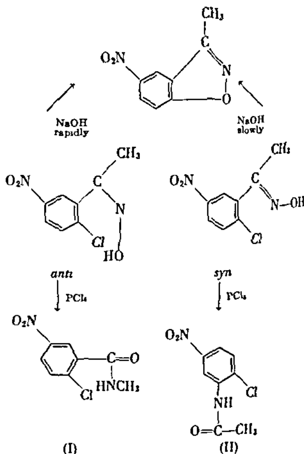
In ketoximes the prefixes syn and anti refer to the relative positions of the hydroxyl group and the group adjacent to the prefix.



Ketoximes undergo the Beckmann rearrangement under the influence of acidic reagents. In this rearrangement, the substituent anti to the hydroxyl group changes positions with the hydroxyl group with the formation of the lactim form of an amide which immediately tautomerizes to the more stable lactam form. Thus, the oxime of acetophenone (*syn*-methyl phenyl ketoxime) yields the lactim form of the stable acetanilide



J. Meisenheimer assigned the presently accepted configurations of the ketoximes largely on the basis of a study of ring-closure reactions involving reactive halogen atoms. For example, one isomer of the oxime of methyl 2-chloro-5-nitrophenyl ketoxime readily undergoes ring closure with elimination of hydrogen chloride under the influence of sodium hydroxide, whereas the other form gives the same product much more slowly. Therefore, it is concluded that the isomer which undergoes facile ring closure is the anti form and that the resistant isomer has the syn configuration. On rearrangement the



anti and syn forms gave (I) and (II) respectively, thus providing a basis for the trans migration of the groups concerned and also providing a basis for the assignment of configuration from the nature of the products of the Beckmann rearrangement.

Cyclohexanone oxime rearranges to the lactam of 6-aminohexanoic acid (caprolactam) a precursor of a polyamide of the nylon type (nylon 66)

—[NH(CH

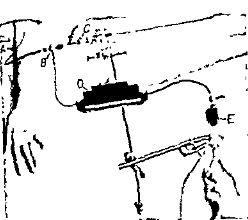


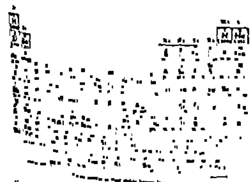
Fig 5. Cuvette oximeter used for immediate and continuous determination of the oxygen saturation of blood being withdrawn from the right heart via a cardiac catheter A, cardiac catheter; B, two-way stopcock, C, a strain-gage manometer, D, cuvette oximeter, E, syringe

ersals of the dye-blood mixture through the circulation following injections of the indicator at various sites in the circulatory system Such curves of concentration with time, indicator dilution curves, are of great value for measurement of blood flow and volume and in the diagnosis of various forms of congenital and acquired cardiac defects. [E.H W]

... photometry of blood in human beings, *Proc Staff Meetings Mayo Clinic*, 25 384–391, 1950, ... H Wood et al, Measurement of blood content and arterial pressure in the human ear, *Proc. Staff Meetings Mayo Clinic*, 25 398–405, 1950, W G ... *A Manual of Reflection Oximetry*, Van Nostrand's Medical Library, no 152, 1958

Oxygen

Chemical element number 8, *oxygen*, O, is one of the gaseous chemical elements Oxygen is of great interest because it is the essential element both in



the respiration process in most living cells and in combustion processes It is the most abundant element in the earth's crust About one-fifth (by volume) of the air is oxygen. Oxygen is separated from air by liquefaction and distillation; the uses which consume the most of this concentrated oxygen are the manufacture, scarfing, cutting, and welding of steel, the mining of taconite, and the production of certain chemicals made on a large scale by partial oxidation processes.

Uncombined oxygen usually exists in the form of diatomic molecules (O_2), but oxygen also exists in a unique triatomic form (O_3) called ozone. *See* OZONE.

In 1774, Joseph Priestley, an English clergyman who emigrated to the United States in 1794 and settled in Northumberland, Pa, observed that mercuric oxide yielded on heating a gas which vigorously supported the combustion of a candle. Priestley found that the gas would *support respiration* and called the gas "dephlogisticated air." The name oxygen, meaning acid former, was proposed for the gas by a group of French chemists in 1787 in recognition of the supposed ability of oxides to form acids

Use in manufacture of chemicals. The largest use for O_2 separated from air is in the manufacture of synthesis gas, hydrogen, acetylene, and other chemicals Hydrogen is produced by removing the oxides of carbon from mixtures of hydrogen and carbon oxides produced by partial oxidation of natural gas Hydrogen, in turn, is used to make ammonia and other chemicals for the hydrogenation-hardening of fats, for use in oxyhydrogen burners and for other purposes. *See* OXIDATION PROCESS

Uses in industry. The second largest use of oxygen is in the iron and steel industry

Mining The extensive deposits of Minnesota taconite, an extremely hard, low-grade iron ore, were not mined on a large scale until a new and economical method was developed for drilling holes preparatory to blasting In this method, burners as high as the depth of the holes to be drilled are mounted vertically above the ground. At the lower end, a very hot flame is produced by burning kerosine in oxygen The heat of the flame spalls the rock The particles thus produced are blown out of the hole by the gases formed in the flame, producing a cloud of steam and dust The burner is lowered continuously as the hole is being drilled. Extremely hard ore can be drilled much faster in this way than by conventional drilling equipment.

Blast furnace In some blast furnaces, oxygen is used to enrich the air supplied to burn the coke. A little extra oxygen increases the daily output appreciably Too much oxygen results in melt... the furnace li... temperature b... oped so that m... oxygen can be used, with consequently greater output per day

Steel from iron O... ver... nac... the open-hearth furnace, oxygen may be injected directly into the melt, here it combines exother-

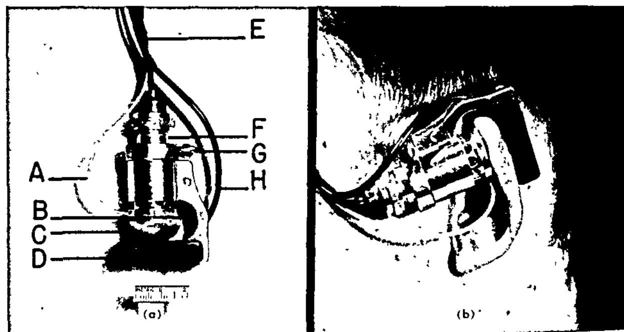


Fig. 3. (a) Earpiece for the oximeter. A, polythene tubing leading into B, pressure chamber, C, rubber diaphragm of pressure capsule; D, housing for photoelectric cells, E, lead wires; F, housing for light source; G, setscrews for fixing position of pressure capsule, H, strain relief and ground wire. (b) Oximeter earpiece in place on ear.

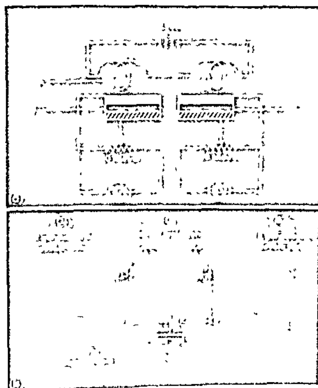


Fig. 4. (a) Diagrammatic arrangement of whole-blood oximeter cell and associated equipment. A, incandescent tungsten-filament light source, B, photocell for red circuit, C, photocell for infrared circuit. (b) Earpiece and cuvette oximeter circuit incorporating simultaneous double-scale and single-scale operation R_1, single-scale potentiometer control for red cell, R_2, single-scale potentiometer control for infrared cell; G_2, single-scale galvanometer (R_1 plus R_2 gives optimal damping resistance); G_3, infrared-cell galvanometer for double-scale operation (R_4 = optimal damping resistance); 1, 2, 3, control-switch positions.

is used to adjust the sensitivity for single-scale operation when the earpiece is on the pressurized (bloodless) ear or a saline solution is in the cuvette oximeter. After adjustments are made at positions 2 and 3, the deflections of the single scale galvanometer produced in switch position 3 are a function of blood oxygen saturation in the flushed ear or cuvette oximeter. An instrument of this type, actually being used for immediate determinations of the oxygen saturation of blood samples being withdrawn via a plastic tube from various sites in the heart and great vessels of a patient undergoing cardiac catheterization, is illustrated in Fig 5.

The cardiac catheter is a woven nylon tube covered with plastic paint containing a tin salt to render it radiopaque. The catheter is inserted via a needle into a vein and advanced under fluoroscopic control into the desired site in the right side of the heart. The two-way stopcock is an interchangeable connection of the cardiac catheter to a strain-gage for recording intracardiac pressures transmitted via the catheter, and to a cuvette oximeter. The determination of oxygen saturation of the blood being withdrawn by the syringe is made in the cuvette oximeter In practice, this blood is maintained sterile and usually is reinfused into the patient after each determination The quantity of blood lost during the procedure is thus minimized.

In addition to many valuable applications requiring intermittent or continuous measurements of the oxygen saturation of arterial or venous blood, oximeters also have been applied to measurement of the blood content of the ear and of arterial blood pressures. More recently, a particularly valuable application has been the use of these devices for continuous measurement of the concentration of indicator dyes in blood the initial trav-

tungsten carbide, a very hard material that imparts great wear resistance to the objects on which it is plated

Most uses for oxygen employ the gas instead of the liquid; even when the oxygen is transported as a liquid for economic reasons, it is usually vaporized before use. However, considerable quantities of liquid are used as oxidizers in the fuel systems of large rockets. See PROPELLANT.

Occurrence. About 49.5% by weight of the earth's crust, including the oceans and atmosphere, is oxygen. Most of this oxygen is combined in the form of silicates, oxides, and water. Water is 88.81% oxygen by weight.

Oxygen also exists outside the earth, but since over 98% of the matter in the visible universe (stars, nebulae, interstellar space) is hydrogen and helium, the cosmic concentration of oxygen is relatively low

Dry air contains 20.946% oxygen by volume, and this concentration has been found to be the same at any level between the surface of the earth and a height of 40 miles. The atoms in atmospheric oxygen consist of three isotopes in the following atomic proportions: 99.759% oxygen-16, 0.037% oxygen-17, and 0.204% oxygen-18. The molecules of oxygen in the air, each of which has two atoms, consist of the statistically expected proportion of the possible combinations of these isotopes, the most abundant molecules being $O^{16}O^{16}$, $O^{16}O^{18}$, and $O^{16}O^{17}$. The isotopic composition of the oxygen in water is slightly different from that in air

Even though large quantities of oxygen from the air are continously being used in respiration, combustion, and other oxidation processes, the concentration of oxygen in the atmosphere remains very nearly constant, chiefly because oxygen is liberated in the process of photosynthesis, in which carbohydrates are produced by green plants from carbon dioxide and water (see PHOTOSYNTHESIS). The primary source of the free oxygen in the atmosphere is believed to have been the decomposition of water vapor by ultraviolet radiation in the upper atmosphere. Almost all the hydrogen formed in this way escapes from the earth's gravitational field, but the oxygen molecules are too heavy to escape, and thus remain in the atmosphere. This photochemical decomposition of water vapor to produce oxygen gas is still going on today

Physical properties. Under ordinary conditions, oxygen is a colorless, odorless, and tasteless gas. It condenses to a pale blue liquid, in contrast to nitrogen, which is quite colorless in the liquid state. Oxygen is one of a small group of slightly paramagnetic gases, and it is the most paramagnetic of the group. Liquid oxygen is also slightly paramagnetic. Some data on oxygen and some properties of the ordinary form, O_2, are listed in the table

Years ago, before the mass spectrometer was invented and when nothing was known about isotopes, the average weight of the oxygen atoms in air was selected by chemists as the standard of weight for all elements and was assigned the value 16.0000. It

Properties of oxygen

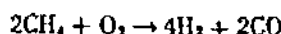
Atomic number	8
Atomic weight	16.0000
Triple point (solid, liquid, and gas in equilibrium)	−218.80°C = 54.36°K
Boiling point at 1 atm pressure	−182.97°C = 90.19°K
Gas density at 0°C and 1 atm pressure, g/liter	1.4290
Liquid density at the normal boiling point, g/ml	1.142
Solubility in water, ml gas per 1000 g water at 25°C and 1 atm pressure	30.8

is now known that isotopes exist and that the isotopic composition of many elements is subject to considerable variation. Consequently, there is no longer a good theoretical basis for the present system of chemical atomic weights. Many chemists now feel that some single isotope (for example, O^{16} = 16.0000 or C^{12} = 12.0000) should be taken as the standard, instead of the mixture of oxygen isotopes as they happen to occur in the earth's atmosphere. See ATOMIC WEIGHT

Chemical properties. Practically all chemical elements except the inert gases form compounds with oxygen. Most elements form oxides when heated in an atmosphere containing oxygen gas. Many elements form more than one oxide, for example, sulfur forms sulfur dioxide (SO_2) and sulfur trioxide (SO_3). Among the most abundant binary compounds of oxygen are water, H_2O, and silica, SiO_2, the latter being the chief ingredient of sand. Among compounds containing more than two elements, the most abundant are the silicates, which constitute most of the rocks and soil. Other widely occurring compounds are calcium carbonate (limestone and marble), calcium sulfate (gypsum), aluminum oxide (bauxite), and the various oxides of iron which are mined as a source of iron. Several other metals are also mined in the form of their oxides. See OXIDE; PEROXIDE

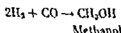
Aside from the sun, man's chief source of energy is the combustion in air, or in more concentrated forms of oxygen, of carbon containing fuels such as coal, petroleum, natural gas, and wood. The principal products of the combustion of these fuels are carbon dioxide, carbon monoxide, and water

Partial oxidation of natural gas is used to make synthesis gas, a mixture of carbon monoxide and hydrogen, by the reaction



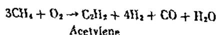
$$2CH_4 + O_2 \rightarrow 4H_2 + 2CO$$

Synthesis gas is used on a large scale to produce various chemicals, such as methanol



$$\underset{\text{Methanol}}{2H_2 + CO \rightarrow CH_3OH}$$

Partial oxidation of natural gas is also used on a large scale for the direct production of valuable chemicals, such as acetylene



$$\underset{\text{Acetylene}}{3CH_4 + O_2 \rightarrow C_2H_2 + 4H_2 + CO + H_2O}$$

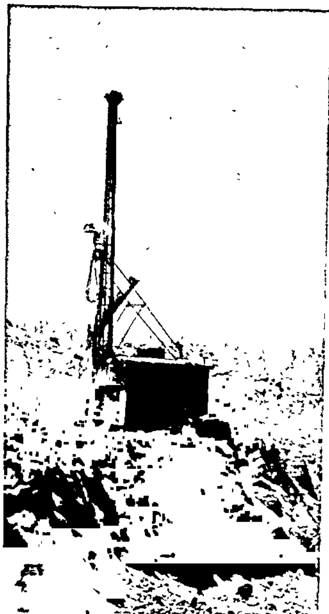


Fig 1. Even the hardest rock spalls quickly under the high heat of jet-piercing machines

mically with the impurities in the iron. It is also possible to use oxygen to enrich the air which supports combustion of the fuels used in the open hearth, thus giving hotter flames which transfer their heat better to the melt. Oxygen may also be injected directly into the melt in electric furnaces. New types of steel converters are being developed to make better use of the potential savings that may be realized by the use of high-concentration oxygen instead of air.

Scarfing. Steel ingots usually contain some surface defects which must be eliminated before the steel can be rolled into high-quality sheet and other products. These defects used to be removed laboriously by grinding or chipping. It is now standard practice in most steel mills to subject the hot ingots, just before rolling, to an oxygen scarfing or skinning process. Streams of oxygen from many nozzles are played on all sides of the ingot at once, thus oxidizing off a thin outer layer of the steel in a great shower of sparks.

Cutting. Although small cuts in steel are still occasionally made with special saws, in most cases oxygen cutting torches are now used. The steel is first warmed at the point where cutting is to start by an oxygen-acetylene flame. A powerful jet of oxygen is then turned on. The oxygen burns some of the iron to iron oxide, and the heat of this combustion melts more iron; the molten iron is blown out of the kerf by the force of the jet. By this process, slabs of steel up to several feet thick can be cut rapidly into desired sizes and shapes. By feeding powdered iron into the oxygen stream, this cutting process can be extended to the cutting of alloys such as stainless steel, which are not readily cut by oxygen alone, and to completely noncombustible materials such as concrete.

Welding. Although a high proportion of welding of steel is now done by one of the available electric arc welding processes, a considerable amount of welding is still done by the older oxyacetylene process. *See* STEEL MANUFACTURE; WELDING AND CUTTING OF METALS.

Other uses. Oxygen is used in the treatment of asthma, pneumonia, and other respiratory diseases. Modern hospitals are piped for oxygen so that compressed oxygen cylinders need no longer be brought into the rooms.

Oxygen is used with hydrogen in the oxyhydrogen burner, which gives a very hot flame and no products of combustion except water. Burners of this type are used in making synthetic sapphire, ruby, and other crystals.

Oxygen is used in flame-plating, a modern process in which metals are covered with protective coatings. A mixture of oxygen, acetylene, and the powder form of the material to be coated, is exploded in a gun, the force of the explosion causes the powder to form an adherent coat on the metal surface. One of the most useful of these coatings is

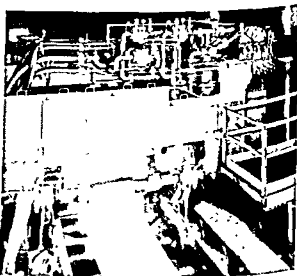


Fig. 2. Continuous hot scarfing takes the place of manual chipping and grinding to remove surface defects in steel products, and gives uniform high-quality surface to steel sheets.

when administered either orally or parenterally (by injection) On oral ingestion, the major site of absorption is the upper part of the small intestine. Considerable amounts of oxytetracycline are removed from the blood by the liver, concentrated in liver tissue, excreted in bile, then resorbed from the intestine into the blood stream. This cyclic process prolongs the period of effective systemic concentration of the antibiotic. Oxytetracycline shows a low degree of toxicity in animals. In clinical use it exhibits a good therapeutic index, that is, there is a wide margin between therapeutic and toxic dosage. Reactions, or side effects, in oxytetracycline treatments are rare, although occasional disturbances of the gastrointestinal tract resulting in nausea or diarrhea have been observed

The reported 1956 United States production estimate for oxytetracycline was about 300,000 lb. *See* ANTIBIOTIC. [C.R.S.]

Bibliography: M. M Musselman, *Terramycin (Oxytetracycline)*, Antibiotic Monographs 6, 1956

Oxyurina

A major group of the Nematoda, also known as the Oxyurata, is one of the two suborders of Ascaridida. In one modern [illegible]

[illegible] component groups such as the Heterakidae, Cosmocercidae and Kathlaniidae are assigned instead to the suborder Ascaridina *See* ASCARIDIDA, OXYUROIDEA. [J.T.L.]

Oxyuroidea

A large and prevalent major group of the class Nematoda For convenience, they are treated here as a superfamily containing the families Oxyuridae, Atractidae, Thelastomatidae, Rhigonematidae, Kathlaniidae, Cosmocercidae, and Heterakidae or Subuluridae About 135 genera have been placed in it. Hosts are terrestrial mammals, birds, reptiles, amphibians, fishes, insects, and some other arthropods

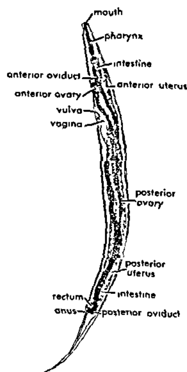
General morphology. The species are small to medium sized, thin bodied Ascaridida There are usually 3 or 6 lips, if present Lateroventral cephalic papillae of the external circle may be present or absent. The buccal capsule or stoma is often small and sometimes rather well-developed Usually the esophagus is grossly divided into a corpus, isthmus, and posterior bulb which often contains a valvular apparatus The excretory system, in most instances is an H-type, with the terminal duct often short and in the form of a reservoir Females are mostly oviparous, some[illegible] reproduc[illegible] times the [illegible] ovoid a[illegible] in most typical genera Males may have 2, 1, or no spicules, and often caudal alae which include some or all of the genital papillae A precloacal sucker may be present in some species Sexual dimorphism is often pronounced

Life cycles. With one exception, known life cycles are direct. Typically the eggs pass out of the host's alimentary tract onto the ground. There they become fully embryonated and infective. Normally the infective egg does not hatch until a susceptible animal ingests it. The cecum and colon of the host are the typical locations of these parasites. Larvae, in all stages of development, as well as adults, occur in the gut. Distribution is cosmopolitan for the group and corresponds to the distribution of the species.

Oxyuriasis. This is the general term for infestations by members of this group, which is less important than several other nematode groups as a cause of disease in man, livestock, and poultry. Among the more common genera are *Enterobius*, *Subulura*, *Probstmayria* and *Heterakis*.

Enterobius vermicularis This oxyurid (see illustration), the human pinworm or seatworm, is common in children. The disease it causes, enterobiasis, is characterized by minor damage to the wall of the lower bowel, pruritis ani, restlessness, and insomnia. Transfer of the eggs from the perianal region to the mouth by the fingers is a common mode of infection

Subulura brumpti. This worm, which inhabits the ceca of chickens and turkeys, is the only oxyuroid known to have an indirect life cycle. In Hawaii, earwigs are its intermediate hosts It is uncommon in the United States and little is known concerning its pathogenicity. Other species of Subulura occur in domesticated birds.



Enterobius vermicularis, adult female (From F. A Brown, *Selected Invertebrate Types*, Wiley, 1950)

Synthesis gas is also a by-product of the acetylene-forming reaction.

Oxygen production. Oxygen is produced on a large scale by the liquefaction and fractional distillation of air. A little oxygen is also made by the electrolysis of water, but oxygen produced this way is more expensive than oxygen from liquid air, and electrolysis of water is not used unless there is some special reason, such as a need for the hydrogen that is also produced

Oxygen for use in chemical laboratories is usually simply withdrawn from steel cylinders of the compressed gas. The traditional methods of preparing oxygen in school chemistry courses are (1) heating potassium chlorate with or without addition of a little manganese dioxide as a catalyst, (2) heating mercuric oxide (Priestley's original method), and (3) electrolysis of water made conducting by addition of sulfuric acid or sodium hydroxide. *See* ATMOSPHERIC GASES, PRODUCTION OF.

Distribution. Oxygen used to be distributed only as gas under pressure in steel cylinders. To avoid the transportation of so great a weight of steel per pound of oxygen, a system for handling liquid oxygen in large, insulated, double-walled tanks has been developed The pressure in the liquid oxygen tanks is only a few pounds per square inch above atmospheric. The insulation is so effective that oxygen-evaporation losses are very small So-called powder-vacuum insulation is used, in which the space between the inner and the outer walls of the tank is filled with insulating powder and the air present is pumped out. At the point of use, the liquid is transferred to stationary storage tanks similarly insulated. When gaseous oxygen is required, liquid is automatically withdrawn from the storage tank and vaporized

There are two important new developments in oxygen distribution, one for the large consumers, the other for the small ones. When more than 100 tons of oxygen is required per day, the oxygen is now usually supplied from a plant built on the consumer's property or through a pipeline from a nearby plant. These on-site plants range in capacity to 1000 tons of oxygen per day.

For the small oxygen consumer, a new kind of liquid oxygen cylinder weighing, when filled, less than 500 lb, of which 250 lb is oxygen, has been developed.

Detection and quantitative analysis. The traditional laboratory test for oxygen gas is that it will cause a glowing wooden splinter to burst into flame; this test does not distinguish between oxygen and nitrous oxide.

In laboratory gas-analysis apparatus, oxygen is usually determined by absorption in an alkaline solution of pyrogallol or in an ammoniacal solution of copper(I) chloride. The concentration of oxygen in oxygen tents and gas streams is readily determined with oxygen meters that measure the content of the oxygen by its paramagnetism. Oxygen in a mixture of gases may be determined in a gas chromatograph. There are a number of colorimetric tests for traces of oxygen. *See* OXIDATION-REDUCTION; RESPIRATION. [C.C.]

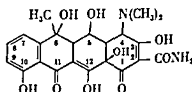
Bibliography: R. F. Benenati, *Oxygen*, in R. E. Kirk and D. F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 9, 1947; J. A. Charles, W. J. B. Chater, and J. L. Harrison, *Oxygen in Iron and Steel Making*, 1956; J. F. Mayberry and T. G. Lutz, Old ways and new, *Welding Engr.*, 39(9):36–38, 1954; F. T. Tancula, Oxygen, its past, present, and future, *Welding Engr.*, 41(11): 44–47, 1956; Upsurge in tonnage O_2 units, *Chem Eng.*, 63(9):354–357, 1956.

Oxytetracycline

A crystalline, amphoteric, broad-spectrum antibiotic elaborated by the actinomycete, *Streptomyces rimosus*, which was found only after the examination of microorganisms from over 100,000 soil samples. The antibiotic is known commercially as Terramycin and is produced by fermentation processes. It is recovered from fermented broths, either through solvent extraction or by precipitation as an insoluble complex salt. In final purification steps, it is crystallized as the hydrochloride salt.

Oxytetracycline is widely used against infectious diseases in both human and veterinary medicine It has broad-spectrum activity against bacteria, spirochetes, rickettsiae, large viruses, and certain protozoa and metazoa, preventing the growth of many organisms at concentrations of less than 1 part per million. It has agricultural application as an animal growth stimulant, in poultry feed to increase egg production, as a food preservative, and as a crop spray ingredient. *See* BACTERIA; RICKETTSIOSES; SPIROCHAETALES; VIRUS

The empirical formula of oxytetracycline is $C_{22}H_{24}N_2O_9$, the molecular weight is 460 and its chemical structure is among the most complicated of natural products, being particularly unusual in the large number of different functional groups present. The stereochemical configuration of the six asymmetric carbon atoms has not yet been reported. The presence of such a large number of



Oxytetracycline

asymmetric centers renders the possibility of an economically feasible chemical synthesis remote Considerable modification of the oxytetracycline structure is possible with retention of antimicrobial activity. Biologically active analogs which vary in substitution at the carbons at positions 4, 5, 6, and 7 are known. *See* CHLORTETRACYCLINE; TETRACYCLINE.

Oxytetracycline is rapidly absorbed and distributed in body tissue in erapeutic concentration

when administered either orally or parenterally (by injection) On oral ingestion, the major site of absorption is the upper part of the small intestine Considerable amounts of oxytetracycline are removed from the blood by the liver, concentrated in liver tissue, excreted in bile, then resorbed from the intestine into the blood stream. This cyclic process prolongs the period of effective systemic concentration of the antibiotic. Oxytetracycline shows a

age. Reactions, or side effects, in oxytetracycline treatments are rare, although occasional disturbances of the gastrointestinal tract resulting in nausea or diarrhea have been observed

The reported 1956 United States production estimate for oxytetracycline was about 300,000 lb. *See* ANTIBIOTIC. [C.R.S.]

Bibliography: M. M. Musselman, *Terramycin (Oxytetracycline)*, Antibiotic Monographs 6, 1956

Oxyurina

A major group of the Nematoda, also known as the Oxyurata, is one of the two suborders of Ascaridida. In one modern system ... essent ... lined ... component groups such as the Heterakidae, Cosmocercidae and Kathlaniidae are assigned instead to the suborder Ascaridina. *See* ASCARIDIDA; OXYUROIDEA [J.T.L.]

Oxyuroidea

A large and prevalent major group of the class Nematoda For convenience, they are treated here as a superfamily containing the families Oxyuridae, Atractidae, Thelastomatidae, Rhigonematidae, Kathlaniidae, Cosmocercidae, and Heterakidae or Subuluridae About 135 genera have been placed in it. Hosts are terrestrial mammals, birds, reptiles, amphibians, fishes, insects, and some other arthropods

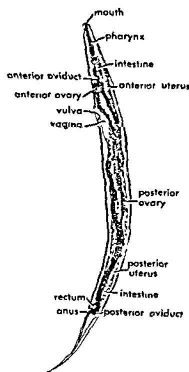
General morphology. The species are small to medium sized, thin-bodied Ascaridida. There are usually 3 or 6 lips, if present Lateroventral cephalic papillae of the external circle may be present or absent The buccal capsule or stoma is often small and sometimes rather well-developed. Usually the esophagus is grossly divided into a corpus, isthmus, and posterior bulb which often contains a valvular apparatus. The excretory system, in most instances is an H-type, with the terminal duct often short and in the form of a reservoir Females are mostly oviparous; sometimes viviparous Their reproductive system is usually didelphic, but sometimes the monodelphic condition occurs Eggs are ovoid and asymmetric in most typical genera Males may have 2, 1, or no spicules, and often caudal alae which include some or all of the genital papillae A precloacal sucker may be present in some species ... ism is often pro-

Life cycles. With one exception, known life cycles are direct. Typically the eggs pass out of the host's alimentary tract onto the ground. There they become fully embryonated and infective. Normally the infective egg does not hatch until a susceptible animal ingests it. The cecum and colon of the host are the typical locations of these parasites Larvae, in all stages of development, as well as adults, occur in the gut. Distribution is cosmopolitan for the group and corresponds to the distribution of the species.

Oxyuriasis. This is the general term for infestations by members of this group, which is less important than several other nematode groups as a cause of disease in man, livestock, and poultry. Among the more common genera are *Enterobius*, *Subulura*, *Probstmayria* and *Heterakis*

Enterobius vermicularis This oxyurid (see illustration), the human pinworm or seatworm, is common in children The disease it causes, enterobiasis, is characterized by minor damage to the wall of the lower bowel, pruritis ani, restlessness, and insomnia Transfer of the eggs from the perianal region to the mouth by the fingers is a common mode of infection

Subulura brumpti. This worm, which inhabits the ceca of chickens and turkeys, is the only oxyuroid known to have an indirect life cycle. In Hawaii, earwigs are its intermediate hosts It is uncommon in the United States and little is known concerning its pathogenicity Other species of *Subulura* occur in domesticated birds.



Enterobius vermicularis, adult female. (From F. A. Brown, Selected Invertebrate Types, Wiley, 1950)

Probstmayria vivipara. This minute atractid pinworm lives in the colon and cecum of the horse; the females produce living young. It reportedly is unique among nematodes of animals in that it can, and regularly does, complete its life cycle entirely within the host's colon (endogenous infection). Horses also commonly harbor *Oxyuris equi* and this large pinworm is somewhat injurious to them.

Heterakis gallinarum. This very common cecal worm of poultry can damage its hosts directly. However, it is more notable for its role in the transmission of blackhead or enterohepatitis, an important disease of turkeys which is caused by a protozoan parasite, *Histomonas meleagridis.* Eggs produced by *Heterakis,* living in the ceca of a fowl also harboring *Histomonas,* often contain *Histomonas.* If a susceptible bird swallows such eggs, after they become infectious on the ground, it can become infected with both cecal worms and the blackhead organism. Other species of *Heterakis* occur in domesticated birds

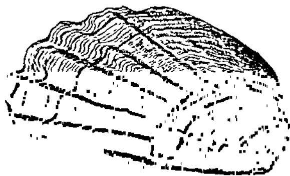
Other species. Oxyuroids do not normally occur in cats, dogs, swine, or cattle. *Skrjabinema ovis* is the characteristic oxyuroid of sheep and goats; it is uncommon in the United States Oxyuroids are common in rodents Among the species that occur in rodents that frequently are used in laboratory experimentation and tests are *Syphacia obvelata* and *Aspicularis tetraptera* of mice, *Heterakis spumosa* of rats, *Passalurus ambiguus* of rabbits, and *Paraspidodera uncinata* of guinea pigs. *See* ASCARIDINA, OXYURINA; PINWORM INFECTION.

[J T L.]

Oyster

Any member of the genus *Ostrea,* class Pelecypoda, phylum Mollusca There are about 100 species in the genus which includes most of the edible oysters. Other related animals are also called oysters, notably the pearl oysters and the spiny tree oysters

Oysters are the most valuable of all marine food animals. The best known species is the common Atlantic, or Virginia, oyster, *Ostrea virginica* It is native to much of the Atlantic Coast and has been widely introduced on the Pacific Coast The West Coast oyster, *O. lurida,* is also commercially important, as is the large Japanese oyster, *O gigantea,* which has been introduced along the Pacific Coast of the United States. There are other species of lesser importance in American waters The European oyster, *O. edulis,* is similar to *O. virginica*



The oyster, *Ostrea virginica;* length to 18 in. (From E. L. Palmer, *Fieldbook of Natural History,* McGraw-Hill, 1949)

Structure. The Virginia oyster may be taken as typical of the group. The shell is irregular, usually broadly tapering, and is large, heavy, and rough. The upper valve is somewhat smaller than the lower. The lower valve is usually attached to some solid object The internal anatomy is essentially similar to that of the related marine clam.

Habitat. Oysters are primarily animals of the bays and estuaries. Wide tolerance in the variation of salinity among marine animals is not common but oysters are especially tolerant of such variations. The Virginia oyster may live in water where the salinity is one-half or even one-third that of the open ocean. Oysters are thus well adapted to the reduced salinity of inlets where the dilution of sea water with fresh water is characteristic. The free-swimming larvae settle to the bottom when they encounter water bearing copper in only slightly greater amounts than sea water, such as is found where there is some dilution of sea water with fresh water.

Oysters feed upon algae, diatoms, and small animals filtered from the sea by cilia on their gills and mantle. *See* ALGAE, DIATOM.

Reproduction. Oysters are among the most prolific of all nonparasitic animals. A large Virginia oyster may produce as many as 500,000,000 eggs in a single season, but 100,000,000 or fewer eggs per individual appears more common. Sexes in *O. virginica* are separate; others, such as *O. lurida,* function as males the first season, females the next This condition is called protandry and animals possessing this cycle are said to be protandric

Eggs and sperm are shed freely into the ocean where fertilization occurs. In 5–6 hours the zygote hatches into a larva of the trochophore type known as a veliger The veliger stage exists only about 48 hours before transformation into a free-swimming bivalve animal. The latter phase lasts for about 2 weeks before the animal settles to the bottom This settling is a critical period for the young oyster; it must come to rest on a solid bottom or perish. If fortunate, it settles on its left valve and becomes firmly attached to its substrate by a secretion produced by the mantle Young attached oysters are called spat The Virginia oyster commonly reaches a length of 5 or more in.; individuals 18 in. long have been collected

Harvesting and use. Harvesting of oysters is most commonly accomplished by dredges or by tonging, grappling with long-handled rakes fastened together scissors fashion and operated by a single fisherman Principal production beds in the United States are the Chesapeake Bay and adjacent areas of the Atlantic Coast.

The meat of the oyster is sold in the shell, fresh shucked, canned, and smoked; the shells are crushed as a source of calcium carbonate for poultry and as a surfacing for ads.

The shell lining of the true oysters lacks the luster commonly associated with bivalve shells and oyster shells are not used for buttons or ornamental mother of pearl Such pearls as may be produced are dull and worthless.

Culturing. Oyster culture has been brought to a high peak of development in Great Britain where the larvae are encouraged to settle on frames and the spat then planted in suitable waters. In American waters spat and oysters too small to utilize are frequently planted in suitable areas. Other than man, the principal natural enemies of the oyster are starfish and a marine snail, the oyster drill.

Pearl oysters. The pearl oysters belong to the family Pteriidae. The shell of these animals is highly iridescent and consequently they may produce pearls of great value. Although pearls of some value are produced off the coasts of Panama and Lower California, the most valuable forms occur off Ceylon and in the Persian Gulf. There is one species of pearl oyster along the Atlantic Coast, the winged pearl oyster, which occurs from North Carolina to the West Indies, but it supports no pearl fishery.

The development of pearl culture by the Japanese is now well established In this intricate process a small bit of foreign material is introduced into the animal between the shell and the mantle. The irritation of this foreign material stimulates the mantle to produce the pearl The Chinese have long practiced a similar art with one of the large fresh water mussels, the rice paddy mussel. They introduce small lead images of Buddha, rather than smaller bits of material, between the mantle and the shell This produces, in a short time, a pearl-coated Buddha, a greatly desired ornament. See CLAM, MOLLUSCA; MUSSEL, OYSTER DRILL; PELECYPODA; STARFISH [J.D.B.]

Oyster drill

A well known snail, *Urosalpinx cinerea*, a member of the family Muricidae, class Gastropoda, phylum Mollusca It ranks with the starfish as one of the two most destructive enemies of the oyster. In some localities, such as Chesapeake Bay, it even exceeds the starfish in damage done to oyster beds There are also other snails that drill oysters. A species of *Thais* is the principal predator in the Gulf area around Pensacola, Florida

The oyster drill is a small snail about 1 or 1¼ in long Its rugged shell is yellowish gray to grayish brown and mottled with varying amounts of white There are five or six broadly shouldered whorls, it is marked with longitudinal ridges, about 10 to each whorl The horny operculum is yellow The outer lip is thin and sharp Its extremely small foot has a yellowish border and is dotted with gray above

The oyster drill settles on a young oyster or other bivalve and, using its strongly toothed radula (rasping tongue), quickly bores a hole in the shell It then sucks the soft parts of the oyster out through this hole

This snail is one of the most common gastropods on the Atlantic Coast, ranging from Prince Edward Island to Florida. It has also become established in San Francisco Bay.

Each female lays 10-100 vase-shaped parchment egg cases each containing about a dozen eggs. These are attached on the underside of rocks or other support, just below the low-water mark. See MOLLUSCA; SNAIL. [J D B.]

Ozokerite

A native mineral wax that occurs near Soldier Summit, at the southwest edge of the Uinta Basin, Utah. The material appears as dark yellow to brown films, veinlets, or nodules disseminated in fracture zones in the shales and sandstones of the Wasatch group of lower Eocene age. The material has a specific gravity of about 0.89, fuses between 60 and 80°C, and is over 99% soluble in carbon disulfide It contains approximately 85% carbon, 14% hydrogen, and 0.3% each of sulfur and nitrogen and is, therefore, nearly a pure hydrocarbon.

Inasmuch as the ozokerite occurs below the oil shales of the Green River formation, it has been thought to be an inspissated (thickened) petroleum derived from sediments of Wasatch age. Chemical structure of ozokerite differs, however, from that of the Wasatch petroleum, leaving some doubt as to this origin

Ozokerite has been recovered by treating the crushed rock in which it occurs with water at 60-70°C. At this temperature the wax melts and floats to the surface Additional purification involves treatment with concentrated sulfuric acid, chromic acid, fuller's earth, or charcoal Such treatment leads to an almost white, higher-melting product known as ceresine, which has been used in the manufacture of candles, shoe polishes, electrical insulation and floor waxes.

Deposits of substances with properties similar to ozokerite also occur in Poland, Romania, Russia, and the Philippines. [I A.B]

Bibliography: H. Abraham, *Asphalts and Allied Substances*, vol. 1, 5th ed , 1945.

Ozone

A high energy allotropic form of the element oxygen Whereas ordinary oxygen has two atoms in each molecule (O_2), ozone (O_3) has three

Ordinary oxygen is a colorless gas and condenses to a very pale blue liquid, whereas ozone gas is decidedly blue, and both liquid and solid ozone are an opaque blue-black color, similar to ink Even at a concentration as low as 4%, the blue color of ozone gas mixed with air or other colorless gas in a tube 1 in or more in diameter and 4 ft or more long can be seen by looking lengthwise through the tube.

Properties and uses. Ozone has a characteristic, pungent odor familiar to most persons because ozone is formed when electrical apparatus produces sparks in air Ozone is irritating to mucous membranes, and toxic to human beings and lower ani-

Probstmayria vivipara. This minute atractid pinworm lives in the colon and cecum of the horse; the females produce living young. It reportedly is unique among nematodes of animals in that it can, and regularly does, complete its life cycle entirely within the host's colon (endogenous infection). Horses also commonly harbor *Oxyuris equi* and this large pinworm is somewhat injurious to them.

Heterakis gallinarum. This very common cecal worm of poultry can damage its hosts directly. However, it is more notable for its role in the transmission of blackhead or enterohepatitis, an important disease of turkeys which is caused by a protozoan parasite, *Histomonas meleagridis.* Eggs produced by *Heterakis,* living in the ceca of a fowl also harboring *Histomonas,* often contain *Histomonas.* If a susceptible bird swallows such eggs, after they become infectious on the ground, it can become infected with both cecal worms and the blackhead organism. Other species of *Heterakis* occur in domesticated birds.

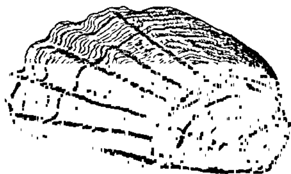
Other species. Oxyuroids do not normally occur in cats, dogs, swine, or cattle *Skrjabinema ovis* is the characteristic oxyuroid of sheep and goats; it is uncommon in the United States. Oxyuroids are common in rodents Among the species that occur in rodents that frequently are used in laboratory experimentation and tests are *Syphacia obvelata* and *Aspicularis tetraptera* of mice, *Heterakis spumosa* of rats, *Passalurus ambiguus* of rabbits, and *Paraspidodera uncinata* of guinea pigs See ASCARIDINA; OXYURINA; PINWORM INFECTION.

[J T.L]

Oyster

Any member of the genus *Ostrea,* class Pelecypoda, phylum Mollusca. There are about 100 species in the genus which includes most of the edible oysters Other related animals are also called oysters, notably the pearl oysters and the spiny tree oysters

Oysters are the most valuable of all marine food animals The best known species is the common Atlantic, or Virginia, oyster, *Ostrea virginica* It is native to much of the Atlantic Coast and has been widely introduced on the Pacific Coast The West Coast oyster, *O. lurida,* is also commercially important, as is the large Japanese oyster, *O gigantea,* which has been introduced along the Pacific Coast of the United States. There are other species of lesser importance in American waters. The European oyster, *O. edulis,* is similar to *O. virginica*



The oyster, *Ostrea virginica,* length to 18 in. (From E. L. Palmer, *Fieldbook of Natural History,* McGraw-Hill, 1949)

Structure. The Virginia oyster may be taken as typical of the group. The shell is irregular, usually broadly tapering, and is large, heavy, and rough The upper valve is somewhat smaller than the lower. The lower valve is usually attached to some solid object. The internal anatomy is essentially similar to that of the related marine clam.

Habitat. Oysters are primarily animals of the bays and estuaries. Wide tolerance in the variation of salinity among marine animals is not common, but oysters are especially tolerant of such variations. The Virginia oyster may live in water where the salinity is one-half or even one-third that of the open ocean. Oysters are thus well adapted to the reduced salinity of inlets where the dilution of sea water with fresh water is characteristic. The free-swimming larvae settle to the bottom when they encounter water bearing copper in only slightly greater amounts than sea water, such as is found where there is some dilution of sea water with fresh water.

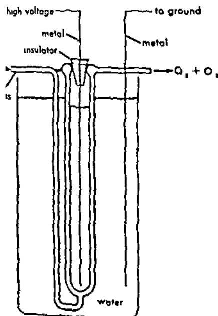
Oysters feed upon algae, diatoms, and small animals filtered from the sea by cilia on their gills and mantle. *See* ALGAE, DIATOM.

Reproduction. Oysters are among the most prolific of all nonparasitic animals. A large Virginia oyster may produce as many as 500,000,000 eggs in a single season, but 100,000 000 or fewer eggs per individual appears more common. Sexes in *O virginica* are separate, others, such as *O lurida,* function as males the first season, females the next This condition is called protandry and animals possessing this cycle are said to be protandric.

Eggs and sperm are shed freely into the ocean where fertilization occurs. In 5-6 hours the zygote hatches into a larva of the trochophore type known as a veliger The veliger stage exists only about 48 hours before transformation into a free-swimming bivalve animal The latter phase lasts for about 2 weeks before the animal settles to the bottom This settling is a critical period for the young oyster; it must come to rest on a solid bottom or perish If fortunate, it settles on its left valve and becomes firmly attached to its substrate by a secretion produced by the mantle Young attached oysters are called spat The Virginia oyster commonly reaches a length of 5 or more in., individuals 18 in long have been collected

Harvesting and use. Harvesting of oysters is most commonly accomplished by dredges or by tonging, grappling with long-handled rakes fastened together scissors-fashion and operated by a single fisherman. Principal production beds in the United States are the Chesapeake Bay and adjacent areas of the Atlantic Coast.

The meat of the oyster is sold in the shell, fresh shucked, canned, and smoked; the shells are crushed as a source of calcium carbonate for poultry and as a surfacing for roads.



:ol silent electric discharge method laboratory generator

these conductors act as electrodes A volt-
:tween 5000–50,000 volts at a frequency be-
50–10,000 cps is then applied across the
ıdes In some commercial ozone generators,
ıer and outer tubes are both water-cooled, in
, only the outer tubes are water-cooled The
represents a simpler type of construction,
ıes not permit as high an input of electrical
as when both tubes are cooled
the plate-type ozonizers, the oxygen or air
ı through a constant-diameter gap of 2–3 mm
en parallel plates, at least one dielectric
must be present in each space between high-
e and grounded surfaces to prevent arcing
ı the electrodes The required voltage is im-
:d on flat electrodes, which may be metal
or metal foil on glass, or electrically con-
ıg paint applied to the glass In constructing
ımercial ozone generator, many plates, sepa-
by spacers, are piled up The plates are usu-
ooled by circulating water in the spaces be-
ı them

concentration of ozone in the gas stream
ıg commercial ozone generators is usually
% by weight The yield of ozone is better when
:n is used instead of air. Other factors which
ase the yield of ozone in the silent electric
arge are thorough drying of the oxygen or air
e it enters the ozonizer, refrigeration, increas-
he pressure to a little above atmospheric, and
asing the frequency of the discharge from 60
0–500 cps

one may be made in the laboratory by the
t electric discharge method, or by the electrol-
of a cold solution of sulfuric acid or perchloric
in water. In the electrolytic method, ozone is
ed as a gas at the anode along with oxygen; the concentration of ozone in the oxygen may be as high as 20% or more by weight Lowering the temperature and increasing the current density at the anode tend to increase the ozone-oxygen ratio in the gas produced.

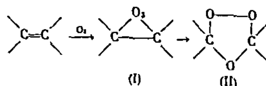
Analytical methods. The analytical determination of ozone is usually carried out in the laboratory by bubbling the gas through a neutral solution of potassium iodide, acidifying the solution, and titrating the iodine thus liberated with standard sodium thiosulfate solution Ozone in a gas stream may be determined automatically and continuously by passing the gas through a cell with transparent windows and measuring the absorption of either visible light or of ultraviolet radiation beamed through the cell *See* OXIDATION-REDUCTION; OXYGEN [C.A C.]

Bibliography: Am Chem. Soc., *Ozone Chemistry and Technology*, Advances in Chemistry, Ser. No. 21, 1959; V. A. Hann and T. C. Manley, *Ozone*, in R. E. Kirk and D F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 9, 1947; C. E. Thorp, *Bibliography of Ozone Technology*, 1954

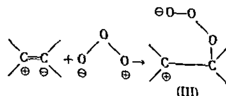
Ozonization

A process using ozone to cleave olefins. It was discovered in 1903 by C. Harries, and it is used to determine the structures of many unsaturated organic compounds, and to prepare a number of other organic compounds

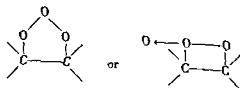
Mechanism. Up to 1950 the reaction of ozone with an olefin was supposed to proceed as follows:



The final true ozonide (II) would yield aldehydes, ketones, or acids upon hydrolysis in the presence of oxidizing or reducing agents Since 1950 it has been postulated that ozone is an electrophilic reagent and that it attacks a polarized double bond in the following manner:



The occasional formation of epoxides, especially with hindered olefins, is best explained by elimination of molecular oxygen from (III). Accordingly, the primary ozonide (I) may exist as:



Some properties of ozone

Density of the gas at 0°C, 1 atm pressure	2 154 g/liter
Density of the liquid	
−111.9°C	1.354 g/ml
−183°C	1.573 g/ml
Boiling point at 1 atm pressure	−111.9°C
Melting point of the solid	−192 5°C
Wavelength range of maximum absorption in visible spectrum	5600–6200 A
Wavelength range of maximum absorption in the ultraviolet spectrum	2400–2800 A

mals. It is not safe to breathe air containing even 0.1 part per million (ppm) of ozone for long periods of time.

Pure ozone and mixtures of ozone with oxygen explode when sparked or otherwise stimulated sufficiently, if the ozone concentration is high enough.

Liquid ozone and liquid oxygen are miscible in all proportions at temperatures above −179.9°C At lower temperatures and in certain composition ranges which vary with the temperature, the liquid mixture separates into two layers, a dense, ozone-rich layer and a lighter, oxygen-rich layer which floats on the heavier one.

Ozone is a more powerful oxidizing agent than oxygen, and oxidation with ozone takes place with evolution of more heat and usually starts at a lower temperature than when oxygen is used.

Ozone sterilizes water more rapidly than chlorine, and readily oxidizes many of the compounds which give contaminated water its bad taste and odor; these properties have led to the use of ozone in the treatment of drinking water in Philadelphia, Pennsylvania, in Paris, France, and in other cities In the presence of water, ozone is a powerful bleaching agent, acting more rapidly than hydrogen peroxide, chlorine, or sulfur dioxide.

An interesting use for ozone, but one which does not require large quantities, is the addition of 1–3 ppm to the air of cold-storage rooms to inhibit bacterial action and mold growth on food

Ozone undergoes a characteristic reaction with unsaturated organic compounds in which the double or triple bond is attacked, even at temperatures as low as −100°C, with the formation of ozonides, these ozonides can be hydrolyzed, oxidized, reduced, or thermally decomposed to a variety of compounds, chiefly aldehydes, ketones, or carboxylic acids. Double (C═C) bonds are almost always ruptured in this reaction. See OZONIZATION.

Ozone is used widely on a laboratory scale in the determination of the structure of organic compounds and in synthetic organic chemistry to cleave double bonds with a minimum of undesirable side reactions. Because many ozonides are explosive, care must be used in carrying out ozonations. Commercially, ozonolysis (ozonation followed by decomposition of the ozonide) is employed in the production of azelaic acid and of various drugs such as cortisone and synthetic sex hormones.

The largest use for ozone is in the production of azelaic acid. More than 1 ton per day of ozone is employed at one plant alone for this purpose. Th first step is the ozonation of oleic acid, known com mercially as red oil and produced from tallow, a by-product of meat-packing plants. The azelaic acid thus produced is esterified to give an importan plasticizer.

Natural occurrence. Ozone occurs to a variabl extent in the earth's atmosphere. Near the earth' surface, the concentration is usually 0 02–0 03 ppm in country air, and less in cities except when ther is smog; under smog conditions in Los Angeles ozone is thought to be formed by the action o sunlight on oxygen of the air in the presence o impurities, and on bad days, the ozone concentra tion may reach figures of 0.5 ppm or more for shor periods of time.

An undesirable effect of ozone in the air is it action on rubber. Even at low concentrations, ozon tends to crack the rubber in automobile tires. Thi effect has been partially overcome by the additio of antioxidants to rubber and by the development o ozone-resistant synthetic rubbers. See ANTIOXI DANT; RUBBER.

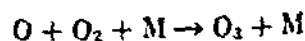
At vertical distances greater than about 13 mile above the earth's surface, ozone is formed by th action of short-wavelength ultraviolet light fro the sun on oxygen of the air. The absolut concentration of ozone is greatest in the strato spheric layer, called the ozone layer, between 1 and 16 miles above the earth's surface. Absorptio of short-wavelength radiation by oxygen to for ozone prevents this radiation from reaching th surface of the earth, and thus helps to make pos sible life on earth, because any appreciable amoun of radiation in this wavelength range would quickl be fatal to man See ATMOSPHERE; IONOSPHERE.

Preparation. The only method used to mak ozone commercially is to pass gaseous oxygen o air through a high-voltage, alternating-current elec tric discharge called a silent electric discharg First, oxygen atoms are formed:



$$O_2 \rightarrow 2O$$

Some of these oxygen atoms then attach themselve to oxygen molecules.



$$O + O_2 + M \rightarrow O_3 + M$$

The excess energy in the newly formed ozone i carried off by any available molecule (M) of ga thus stabilizing the ozone molecule

Ozone generators are of two types, the concen tric-tube and the plate type. In the concentric-tub type, the oxygen or air to be ozonized passe through the annular space (about 2–3 mm across between two tubes, one of which must be made of dielectric material, usually glass, and the other o which may be either glass or a metal which doe not catalyze ozone decomposition, such as alumi num or stainless steel. The internal surface of th inner tube and the external surface of the oute tube, when made of glass, are in contact with a electrical conductor such as metal foil, an elec trically conducting paint, or electrically conductin

P Pacific Islands to Peptide

Pacific islands

The islands of the Pacific Ocean are sparse in the north and east and common in groups through the central and particularly the southwestern Pacific regions. On most maps of the whole Pacific Ocean, by far the greater proportion of the islands are smaller than an observable mark if drawn to scale. Only the outstanding islands and groups are discussed in this article. For New Guinea and the Bismark Archipelago, see EAST INDIES.

Hawaiian Islands The Hawaiian Islands are a chain of islands, reefs, and shoals stretching northwest southeast for nearly 2000 miles. The islands are great dome volcanoes that have been built up from the floor of the ocean along a line which runs from Kure Island to a point southeast of Hawaii where the volcanoes have not yet reached the surface of the ocean. The total land area is 6441 square miles. Over 99% of this is contained in the southeastern group of eight high, volcanic islands: Hawaii, 4030 sq mi; Kahoolawe, 45 sq mi; Maui, 728 sq mi; Lanai, 141 sq mi; Molokai, 260 sq mi; Oahu, 604 sq mi; Kauai, 555 sq mi; and Niihau, 72 sq mi. The middle group from Kaula to Gardner consists of pinnacles or low mountain tops. The northwestern group from Laysan to Kure consists almost entirely of reefs, atolls, banks, and shoals with no volcanic rock visible above the sea.

The volcanoes probably appeared above sea level in the middle or late Tertiary. The weight of the material appears to have depressed the floor of the ocean to a depth of 18,000 ft. Mauna Kea rises almost 32,000 ft from its base to its summit 13,784 ft above sea level. The western islands are older and more eroded. Distinctive features of the is-

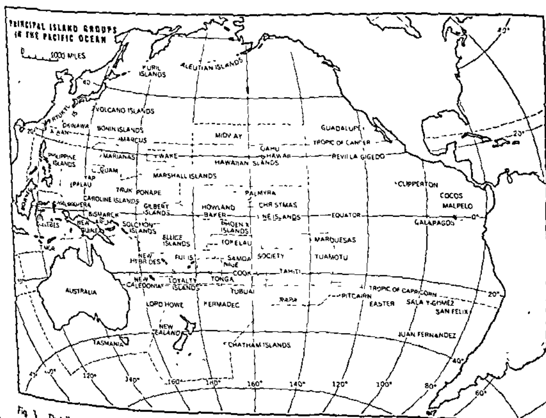
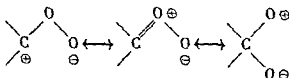
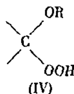


Fig. 1. Distribution of the principal island groups of the Pacific Ocean. (From O. W. Freeman and J. W. Morris, World Geography, McGraw-Hill, 1958)

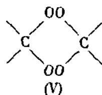
which decomposes to give an aldehyde or ketone and a zwitterion which may have the following resonance structures:



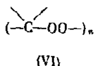
The zwitterion may react with an ionic solvent such as water, alcohol, or acid to give an oxyperoxide (IV) where R is H, an alkyl group, or an acyl group; or it may react with itself to give a peroxide dimer (V) or polymeric peroxide (VI). In nonionizing solvents such as carbon tetrachloride, the zwitterion may react with the aldehyde or ketone to give the classical ozonide (II), or it may form a polymeric ozonide (VII).



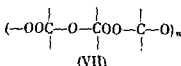
(IV)



(V)



(VI)



(VII)

Technique of ozonization. Ozone is readily produced from air or oxygen by passage of a high-voltage electrical discharge through a stream of the gas. Concentrations of ozone for optimum electrical efficiency are 1 wt% for air and 2 wt% for oxygen at pressures of 2–8 lb/psig. A solvent is usually employed for the unsaturated compound during ozonization. In reactive solvents, such as methanol or acetic acid, a methoxy or acetoxy hydroperoxide is formed as an intermediate. In nonreactive solvents, such as chloroform or carbon tetrachloride, a polymeric peroxide or ozonide may be the intermediate. Decomposition of the intermediate to useful products may be induced under reducing conditions, producing an aldehyde or alcohol; or under oxidizing conditions, using hydrogen peroxide, air, oxygen, or slightly ozonized oxygen, producing an acid as the final product.

The location of the unsaturated linkages can be determined by identifying the fragments resulting from the decomposition of the ozonides. For instance, the isolation of propionaldehyde CH_3CH_2CHO, and diethyl ketone, $CH_3CH_2COCH_2CH_3$, from the ozonization of an alkene identifies the original alkene as the compound *3-ethyl 3-hexene*, $(CH_3{-}CH_2{-})_2C{=}CH{-}CH_2{-}CH_3$.

Industrial utilization. The commercial production of azelaic and pelargonic acids by the ozonolysis of oleic acid represents the first large-scale use of ozone for the synthesis of industrially important chemicals. Oleic acid, dissolved in about half its weight of pelargonic acid obtained from a previous batch is ozonized continuously in a countercurrent reactor with about 2% ozone in oxygen. The ozonized mixture is oxidized continuously with a stream of oxygen over a period of several hours until the active oxygen content is reduced to a minimum value. The pelargonic and azelaic acids are recovered from the mixture by vacuum distillation.

The synthesis of ω-aminopelargonic acid by the ozonolysis of oleoyl nitrile and oleoyl amine has been extensively investigated both in England and Japan. Production of a nylon-9 fiber from the lactam is reported from the latter country. *See* UNSATURATED HYDROCARBON. [C.A.C.]

Bibliography: P. J. Bailey, *Chem. Revs.*, 1958.

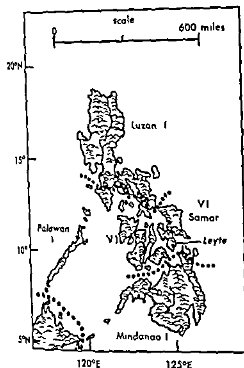


Fig 3 Sketch map of principal physical regions and land surface character of the Philippine Islands VI indicates the Visayan Islands region, B designates adjacent northeast Borneo. (*Modified from a base map by Erwin Raisz in G. B Cressey, Asia's Lands and Peoples* 2d ed , McGraw-Hill, 1951)

plexity of the structure is shown by the fact the channel between Panay and Negros is 110 ...ep while the channel between Negros and Cebu ...000 ft deep Mixtures of coral and volcanic ...erials commonly form soils which are unusually ...for the tropics

...Mindanao Although Mindanao is almost as large ...Luzon, it is much more rugged and lacks low...ds. The Cotabato and Agusan plains are the re... of recent silting of bays Between the two de...ssions is the active volcano Mt Apo, 9690 ft ...ke Lanao at an elevation of 2200 ft is the result ...a lava flow which dammed a valley.

Palawan Palawan is a narrow island which is ...0 miles long with mountains rising 6600 ft above ...e sea. In the south it is composed of crystalline ...ck and in the north of limestone [C A M A]

Bibliography: O W. Freeman (ed), *Geography of the Pacific*, 1951, C. Robequain, *Malaya, Indonesia, Borneo, and the Philippines*, 2d ed., ...958, H T. Stearns, *Geology of the Hawaiian Islands*, Territ Hawaii, Div Hydrog Bull. 8, 1946

Pacific Ocean

The Pacific Ocean has an area of 165,000,000 square kilometers and a mean depth of 4282 km ...t covers 32% of the earth's surface and 46% of the surface of all oceans and seas, and its area is greater than that of all the land areas combined. Its mean depth is the greatest of the three oceans and its volume is 53% of the total of all oceans. Its greatest depths in the Marianas and Japan trenches are the world's deepest, more than 10 km (Fig. 1).

Surface currents. The two major wind systems driving the waters of the ocean are the westerlies which lie about 40-50° latitude in both hemispheres (the "roaring forties") and the trade winds from the east which dominate in the region between 20°N and 20°S These give momentum directly to the west wind drift (flow to the east) in high latitudes and to the equatorial currents which flow to the west. At the continents there is flow of water from one system to the other and huge circulatory systems result (Fig. 2).

The swiftest flow (greater than 2 knots) is found in the Kuroshio Current near Japan. It forms the northwestern part of a huge clockwise gyral whose north edge lies in the west wind drift centered at about 40°N, whose eastern part is the south flowing California Current, and whose southern part is the North Equatorial Current

Part of the west wind drift turns northward into the Gulf of Alaska, thence westward again and into the Bering Sea, returning southward off Kamchatka and northern Japan where it is called the Oyashio Current

H. U. Sverdrup (1942) has reported estimates of the flow in the North Pacific in the upper 1500 meters The Kuroshio carries about 65,000,000 m³/sec at its greatest strength off Japan The west wind drift in mid-ocean carries about 35,000,000, the California Current east of 135°W carries about 15,000,000, and the North Equatorial *Countercurrent* about 45,000,000

A gyral corresponding to the Kuroshio-California-North Equatorial current gyral is found in the Southern Hemisphere Its rotation is counterclockwise, with the highest speeds (about *2 knots*) in the Southeast Australia Current at about 30°S. The current turns eastward and flows around New Zealand to South America, where it turns northward Along this coast it has *been called the* Humboldt or the Peru Current It turns westward at the Equator and is known as the South Equatorial Cur-...

...torial edge of the gyral is not so regular in shape nor so *clearly* defined as its northern counterpart. See SOUTHEAST ASIAN WATERS.

In the region of the west wind drift in the South Pacific the ocean is open both to the *Indian* and the Atlantic, although the eastward passage to the Atlantic through Drake Passage is narrower and shallower *than the* region between Australia and Antarctica Part of the water flows, however, all around Antarctica with the wind behind it, and it *receives more* momentum than its northern counter-

lands are steep cliffs, amphitheater-headed valleys, and drowned valleys.

The climate is determined by location within the tropics in the midst of the world's largest ocean. At sea level there is no frost, and the variation between the mean monthly temperatures is less than the diurnal range. Temperatures average about 72°F; extremes of hot and cold are lacking. Temperatures decrease with elevation, and frost occurs above 4000 ft. Snow falls in winter on the highest peaks, Mauna Kea and Mauna Loa on Hawaii, and Haleakala on Maui A high relative humidity is modified by the prevailing northeast trade winds. Ordinarily the trade winds are dominant except for a short period in August and September.

Rainfall varies according to seasons, elevation, and exposure. The winters are rainy and the summers dry. On the windward slopes which are exposed to the northeast trade winds the annual rainfall at elevations above 2500 ft may be as high as 200 or 300 in. Near the top of Mt Waialeale on the island of Kauai, 624 in. of rainfall has been recorded in a single year. In contrast, lowlands on the leeward sides of the mountains receive an average annual rainfall between 4 and 20 in.

Hawaii, the biggest and most recent of these islands, has an area greater than that of Delaware and Rhode Island combined It is the only island which possesses active volcanoes. Mauna Loa which erupts on an average of once every 3–4 years and Kilauea whose last eruptions were in 1952, 1954, 1955, and 1959

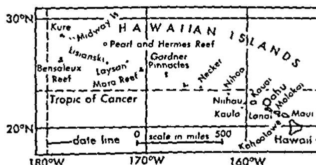


Fig. 2 Location and extent of the Hawaiian Islands

New Caledonia. New Caledonia, or Nouvelle Calédonie, is a mountainous island 248 miles long and 30 miles wide with an area of 6200 square miles. Its dependencies, the Isle of Pines and the Loyalty Islands, add another 800 square miles The island is possibly an extension of the folded Owen Stanley Mountains of New Guinea. It is composed of two parallel mountain ranges with elevations up to 5000 ft, a broad interior plateau, and small coastal plains. The ancient crystalline rocks are rich in minerals, and New Caledonia has been at times a major world source of nickel and chromite. Other minerals have been exploited to a lesser extent but appear to be important: iron, manganese, coal, gold, silver, lead, and zinc.

A short rainy season from January through March is followed by light rainfall the rest of the year, with an average annual rainfall of about 40 in. in the lowlands. The mean monthly temperature ranges from 65°F in July to 72°F in December Heavier rains in the interior result in a tropical scrub forest vegetation while the drier plains have a cover of savanna and brush.

The Philippines. The Philippine Islands, a group of over 7000 islands stretching a distance of 1200 miles, have a total land area of 115,000 square miles. However, only 31 islands have an area of more than 100 square miles each. The three largest are Luzon, 40,400 sq mi; Mindanao, 36,520 sq mi, and Samar, 5,049 sq mi. The present islands are of late Tertiary and Quaternary origin but ancient rock of Precambrian and Mesozoic times indicates land has been in this area a long time and has had a varied history. The basic structure is a fold which descends abruptly on the east into the Philippine Deep, which is one of the greatest openings in the surface of the earth. The floor of the Philippine Deep at its maximum is 7 miles below the surface of the ocean. The varied geological history of the islands has resulted in important deposits of minerals, although little has been done to exploit them High-grade, often extensive iron deposits are found on all of the main eastern islands; coal occurs on Cebu; manganese occurs in the Visayan Islands; and chromite and gold are found on Luzon.

The low latitude and warm ocean currents combine to give the islands an even, high temperature with a range from 75 to 85°F Exceptions to this occur when occasional cool air masses push over from the continent during the winter and bring temperatures as low as 65°F. The greatest climatic differences are due to variations in rainfall Differences in topography result in great rainfall variation within short distances In general the east coast, which receives typhoons, is wetter than the west coast, which has a pronounced dry season (see HURRICANE) However, there are exceptions, such as Masbate which has an annual average rainfall of 72 in, while Baguio in the mountains of Luzon has 179 in Tropical rainforest conditions once prevailed over most of the islands and some remain, especially on Luzon and Mindanao.

Luzon. The largest and most massive of the islands, Luzon, is characterized by a north south trend of its mountains and rivers The Central Mountains have several peaks over 6000 ft, the highest of which is Mt. Pulog, 9600 ft The Central Basin and the Cagayan Basin are graben which have been filled and offer the most extensive lowlands of the Philippines. Southern Luzon is deeply indented with bays and gulfs. With the exception of Mt Mayon, the numerous volcanoes of southern Luzon are mostly extinct or dormant.

Visayan Islands. The islands of the central Philippines which surround the Visayan Sea are grouped together as the Visayan Islands These include Samar, Negros, Panay, Leyte, Cebu, Bohol and Masbate. Although these islands owe much to volcanic activity which is still going on, they also contain pre-Tertiary rocks, including granite The

this level while the upper waters (South Equatorial Current) were flowing west at 0 5–1 5 knots It has been suggested that this equatorial undercurrent be called the Cromwell Current, after its discoverer. See OCEAN CURRENTS.

Temperature at the sea surface. Equatorward of 30° latitude the heat received from the sun exceeds that lost by reflection and back radiation, and rface waters flowing into these latitudes from gher latitudes (California and Peru Currents) increase in temperature as they flow equatorward and ra west with the Equatorial Current System ey carry heat poleward in the Kuroshio and utheast Australia Currents and transfer part of [illegible]

juatorward currents along the *eastern* boundaries f the subtropical anticyclones is thus much wer than that of the currents of their western undaries at the same latitudes. Heat is accumulated, and the highest temperatures (more than 8°C) are found at the *western end* of the equatorial region (Fig. 3). Along the Equator itself somewhat lower temperatures are found. The cold Peru Current contributes to its eastern end, and there is apparent upwelling of deeper, colder water at the Equator, especially at its *eastern end* (as low as 19°C in February at 90°W) as a result of the divergence in the wind field

Upwelling also occurs at the edge of the eastern boundary currents of the subtropical anticyclones When the *winds blow strongly equatorward* (in summer) the surface waters are driven offshore, and the deeper colder waters rise to the surface and further reduce the low *temperatures of these equatorward flowing currents.* The effect of these seasonal variations in the winds is thus to reduce the seasonal range of temperature, *since the upwelling* occurs in spring and summer. The seasonal range of nearshore surface temperature of the California Current at 35°N is less than 4C° (9–13°C), *though* the latitudinal mean is about 10°C. The equatorward winds off Japan occur in winter and the poleward in summer, so that seasonal range is increased at that latitude to more than 16C° (10–26°C).

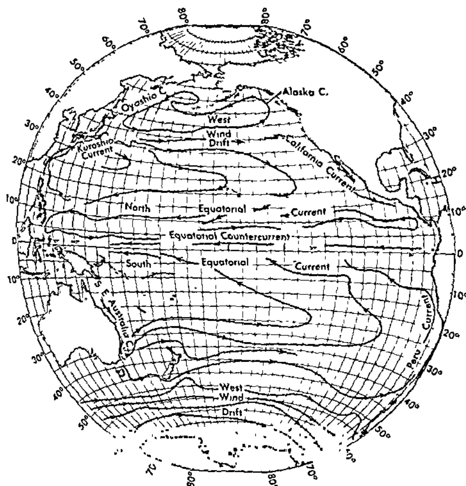


Fig 2 The principal currents of the Pacific Ocean Prepared from computations of geostrophic flow at the sea surface relative to 1000 decibars, using data from the NORPAC, EQUAPAC, Carnegie, and Discovery expeditions

part. The total transport is several times greater. G. E. R. Deacon (1937) has estimated the transport to the east in the South Pacific part of the west wind drift as more than 100,000,000 m³/sec in the upper 3000 m. Sverdrup estimates only about 35,000,000 m³/sec for the upper 1500 m in the North Pacific west wind drift. The gyral corresponding to the Oyashio-Gulf of Alaska gyral of the North Pacific is thus vaster in transport and area, since much of it passes around Antarctica. *See* ANTARCTIC OCEAN; INDIAN OCEAN.

Between the two subtropical anticyclones (Kuroshio-California-North Equatorial Current and the Southeast Australia-Peru-South Equatorial Current) lies an east-flowing current between about 5° and 10°N, called the Equatorial Countercurrent. Sverdrup has estimates of flow of 25,000,000 m³/sec for this current. Some of the observed set and drift of vessels in the area suggest that there is also an eastward flow along 10°S from 155°E to 180° from November through April.

Within the upper 1000 m the flow of water is generally parallel to the surface flow, but slower. Certain important exceptions occur. Beneath the surface waters of the California and Peru Currents, at depths of 200 m and below, countercurrents have been found, in which some part of the tropical waters are carried poleward. In the California Current this flow reaches to the surface in December, January, and February, and it is known as the Davidson Current north of 35°N. It is not known whether a surface poleward flow occurs in southern winter in the Peru Current.

Recent direct measurements at the Equator and 140°W have revealed that a subsurface flow to the east is found at least from 140°W to 90°W, with highest velocities at depths of about 100 m. Speeds as high as 2–3.5 knots to the east were found at

Fig. 1. The principal relief features of the Pacific Ocean. Trenches slightly exaggerated in scale. (*From F. P. Shepard, The Earth Beneath the Sea, Johns Hopkins, 1959*)

wind drift, contribute water of low salinity to the eastern boundary currents (off California and South America). The greater part of the effect of the eastern tropical precipitation is found at the surface of the North Equatorial Current and Countercurrent. Near Central America the values are less than 33 ‰ at 10°N, but they rise nearly to 34.5 near the Philippine Islands.

Dissolved oxygen at the sea surface. Above the thermocline the water is in continual overturn and is thus in constant contact with the atmosphere. Oxygen from the atmosphere dissolves in the water until equilibrium is established, and over most of the Pacific the upper layer is very close to saturation in oxygen content with typical values from about 98 to 103% of the saturation value. *See* THERMOCLINE.

The saturated value of dissolved oxygen rises as both the temperature and salinity fall, but the range of surface temperature in the ocean accounts for a wider variation in saturated value than does that of surface salinity, and it is principally the variation in space and time of surface temperature which accounts for the oxygen values at the surface. Values greater than 7 ml/liter are

[illegible]

inorganic phosphate-phosphorus, silicate-silicon, and nitrate generally increase from the sea surface downward, since photosynthesis and growth in the upper mixed layer tend to use such quantities as are there, and diffusion upward from the higher concentrations is limited by the great stability usually found immediately below the surface layer. At the surface phosphate-phosphorus varies from less than 0.25 microgram-atoms (μg-atoms) per liter in the centers of the anticyclones to more than 1.5 in the high latitude cyclones. High values are also found in the California and Peru Currents. In a manner similar to the low temperatures found there, their high concentrations are partly the result of transport of mixed water from the cyclones and partly the result of upwelling at the coasts under equatorward winds. Values greater than 1 μg-atom/liter are found in both areas during the summer period of upwelling. At the Equator in the eastern Pacific upwelling raises the values to more than 1 throughout the year.

Silicate-silicon has not been so extensively measured. It also is low in value at the surface and increases with depth. Surface values range from as high as 40 μg-atoms/liter in the high-latitude cyclones and 12 in the upwelling regions of the California Current, to 4 or less in the center of the anticyclones and to values too small to be detected near the Equator in the eastern region.

Distribution of properties with depth. The surface waters in high latitudes are colder and heavier than those in low latitudes. As a result some of the high latitude waters sink below the surface and spread equatorward, mixing mostly with water of their own density as they move and eventually providing the dominant water type (in terms of salinity and temperature) of that density over vast regions (Fig. 4).

The deep and bottom water of all oceans is believed to be formed in the Atlantic high latitudes

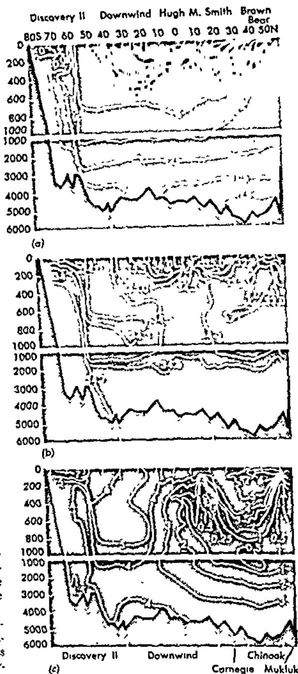


Fig. 4 Vertical sections of (a) *temperature* in degrees C, (b) salinity in parts per thousand, and (c) dissolved oxygen in milliliters per liter in the central *Pacific* Ocean, approximately along the meridian 160°W, from Alaska (right) to Antarctica (left). Depths are in meters. Depth scale is expanded in upper 1000 meters. Data from Carnegie expedition, NORPAC and EQUAPAC expeditions, Discovery expeditions, and Chinook, Mukluk, and Downwind expeditions

The temperatures at the surface of the South Pacific Ocean have not been nearly so well documented, since most of the information comes from measurements made by merchant vessels, and the commercial shipping lines cover but a small part of its great extent. It may be reasoned that most of the temperature characteristics of the North Pacific will have analogies in the Southern Hemisphere. The damped seasonal variation of the California Current seems to occur in the Peru Current as well. But the temperatures of the Southeast Australia Current do not seem to vary so widely through the year as those of the Kuroshio, probably because of the restrictions upon the flow which are imposed by the islands.

The limiting temperature in high latitudes is that of freezing. Ice is formed at the surface at temperatures slightly less than −1°C depending upon the salinity; further loss of heat is retarded by its insulating effect. The ice field covers the northern and eastern parts of the Bering Sea in winter, and most of the Sea of Okhotsk, including that part adjacent to Hokkaido (the north island of Japan). Summer temperatures, however, reach as high as 6°C in the northern Bering Sea and as high as 10° in the northern part of the Sea of Okhotsk.

Pack ice reaches to about 62°S from Antarctica in October and to about 70°S in March, with icebergs reaching as far as 50°S. See BERING SEA; ICEBERG; SEA ICE.

Salinity at the sea surface. The highest values of salinity observed in the Pacific Ocean are slightly greater than 35.5 and 36.5 parts per thousand (‰), found respectively in the surface water of the centers of the north and south subtropical anticyclones. These anticyclones cover the latitudes in which evaporation exceeds precipitation, and the overlying anticyclonic winds oppose outward flow at the sea surface. The three regions where precipitation most strongly exceeds evaporation are found poleward of 40° latitude and in the eastern tropical Pacific. The result is that the high-latitude cyclones are regions of low salinity (as low as 32.5 ‰ in the north, 33.8 in the south) which, through mixing with the anticyclones in the region of the west

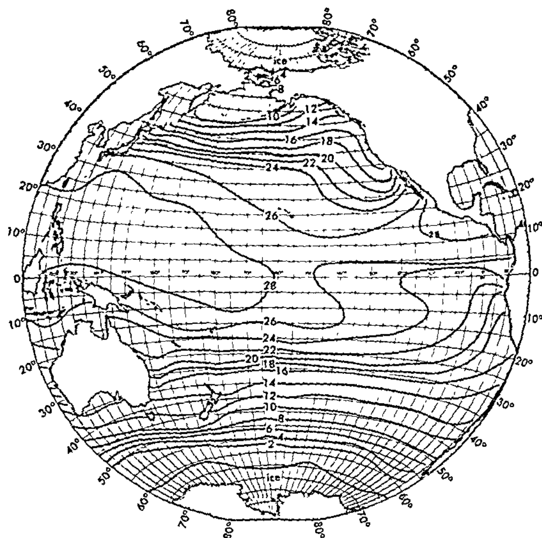


Fig. 3. Temperature at the sea surface in August in degrees C. (Adapted from U.S. Navy Hydrographic Office, H. O. Publ. 225, 1948)

wind drift, contribute water of low salinity to the eastern boundary currents (off California and South America). The greater part of the effect of the eastern tropical precipitation is found at the surface of the North Equatorial Current and Countercurrent. Near Central America the values are less than 33 ‰ at 10°N, but they rise nearly to 34.5 near the Philippine Islands.

Dissolved oxygen at the sea surface. Above the thermocline the water is in continual overturn

he Pacific the upper layer is very close to saturation in oxygen content with typical values from bout 98 to 103% of the saturation value *See* THERMOCLINE.

The saturated value of dissolved oxygen rises as both the temperature and salinity fall, but the range of surface temperature in the ocean accounts for a wider variation in saturated value than does that of surface salinity, and it is principally the variation in space and time of surface temperature which accounts for the oxygen values

inorganic phosphate-phosphorus, silicate-silicon, and nitrate generally increase from the sea surface downward, since photosynthesis and growth in the upper mixed layer tend to use such quantities as are there, and diffusion upward from the higher concentrations is limited by the great stability usually found immediately below the surface layer. At the surface phosphate-phosphorus varies from less than 0.25 microgram-atoms (μg-atoms) per liter in the centers of the anticyclones to more than 1.5 in the high latitude cyclones. High values are also found in the California and Peru Currents. In a manner similar to the low temperatures found there, their high concentrations are partly the result of transport of mixed water from the cyclones and partly the result of upwelling at the coasts un-

Eastern Pacific upwelling raises the values to more than 1 throughout the year.

Silicate-silicon has not been so extensively measured. It also is low in value at the surface and increases with depth. Surface values range from as high as 40 μg atoms/liter in the high-latitude cyclones and 12 in the upwelling regions of the California Current, to 4 or less in the center of the anticyclones

in low latitudes. As a result some of the high latitude waters sink below the surface and spread equatorward, mixing mostly with water of their own density as they move and eventually providing the dominant water type (in terms of salinity and temperature) of that density over vast regions (Fig 4).

The deep and bottom water of all oceans is believed to be formed in the Atlantic high latitudes

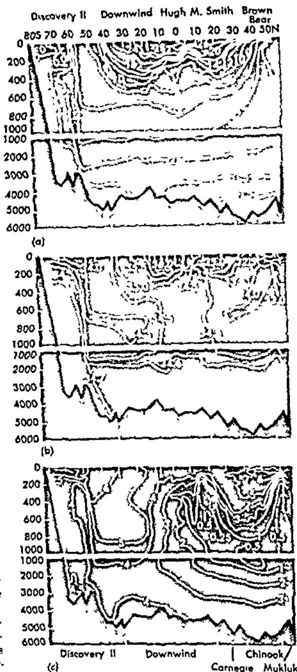


Fig 4. Vertical sections of (a) temperature in degrees C, (b) salinity in parts per thousand, and (c) dissolved oxygen in milliliters per liter in the central Pacific Ocean, approximately along the meridian 160°W, from Alaska (right) to Antarctica (left). Depths are in meters. Depth scale is expanded in upper 1000 meters. Data from Carnegie expedition, NORPAC and EQUAPAC expeditions, Discovery expeditions, and Chinook, Mukluk, and Downwind expeditions.

off Greenland and in the Weddell Sea. It arrives in the South Pacific with temperature less than 2°C, salinity about 34.65–34.7‰, and oxygen about 4.5 ml/liter. The waters filling the Pacific below 3000 m retain these values of temperature and salinity everywhere, but in the northern part the oxygen is reduced to values between 2.5 and 3.5. This would be consistent with depletion by decay and respiration during a slow movement to the north.

Over most of the North Pacific Ocean the temperature does not decrease all the way to the bottom, but beneath a minimum value at about 3800–4000 m it rises again. The increase in temperature is probably in close balance with that in pressure, since no evidence has been found of instability. Two possible explanations for the temperature maximum at the bottom are the flow of heat upward from the ocean floor and an adiabatic rise in temperature as the water flows downward into the deeper basins.

The most conspicuous water masses formed in the Pacific are the Intermediate Waters of the North and of the South Pacific, which on the vertical sections include the two huge tongues of low salinity extending equatorward beneath the surface from about 55°S and from about 45°N. The southern tongue is higher in salinity and density and lies at a greater depth, since the surface waters of the high-latitude cyclone are more saline in the south than in the north.

It has been observed that the higher salinities are found at the surface in the anticyclones. The highest values are in the equatorward halves of the anticyclones. High values penetrate the base of the mixed layer, and tongues of high salinity extend in the thermocline toward the Equator from both north and south.

Beneath the mixed layer the waters are not in contact with the atmosphere and the oxygen consumed cannot be replaced directly from the atmosphere. The oxygen quickly falls below the saturated value. Even where high values of oxygen accompany the sinking of water masses, as in the South Pacific Intermediate Water, the oxygen is not in saturation below 300 m.

Between the saturated waters of the surface and the cold bottom waters entering from the south lies a minimum value of oxygen. Beneath the tongue of high oxygen associated with the South Pacific Intermediate Water the minimum is only a little less than 4 ml/liter. Beneath the North Pacific Intermediate Water, however, no water as high as 3.5 is found, and in the minimum itself values less than 0.5 occur over large areas. The values of oxygen beneath the surface are the result of consumption by organisms and replenishment by mixing and renewal of the water. Since at any position the values are nearly constant in time, consumption must everywhere equal replenishment by both flow and diffusion.

Phosphate-phosphorus increases rapidly beneath the sea surface to a maximum which usually lies beneath the oxygen minimum. In regions of upwelling and divergence, higher values of phosphate and other nutrients from depth may from time to time be brought to the surface and thus made available to the plants. Such regions (California and Peru Currents, South Equatorial Current) are highly productive. The values gradually diminish beneath the maximum, to about 2.5 μg-atoms/liter at the bottom in the north and slightly less than 2 in the south. The maximum value is greater than 3.5 in the north and between 2 and 2.5 in the south.

It is to be noted that the Pacific Ocean is higher in phosphate concentration than the Atlantic or Indian oceans, exceeding the Atlantic by about 1 μg-atom/liter on the average, though the values at the surface do not differ so much. This excess, like the lower value of oxygen, is undoubtedly related to the deep circulation.

Nitrate-nitrogen is present in the ratio of approximately 8:1 by weight of phosphate-phosphorus in those areas where it has been measured, but otherwise the details of its distribution are not so well known.

Silicate-silicon has also not been adequately sampled. Its vertical distribution parallels that of phosphate-phosphorus except that beneath the level of the phosphate maximum it remains nearly constant, at about 170 μg-atoms/liter in the center of the northern anticyclone, as high as 220 in the Bering Sea, about 130 in the center of the southern anticyclone. Silicate-silicon, like phosphate-phosphorus, is more highly concentrated in the Pacific than in the Indian and Atlantic oceans. *See* MARINE SEDIMENTS, OCEANOGRAPHY, SEA WATER; SEA WATER FERTILITY, SUBMARINE TOPOGRAPHY.

[J.L.R.]

Bibliography: G. E. R. Deacon, *Discovery Reports*, vol. 15, 1937; *Discovery Investigations Station List 1931–1933*, vol. 21, 1941; J. A. Fleming et al., *Scientific Results of Cruise VII of the Carnegie during 1928–29*, Carnegie Inst. Washington Publ. 545, Oceanog. vol. I-B, 1945; E. Kossinna, *Die Tiefen des Weltmeeres*, 1921; NORPAC committee, *Oceanic Observations of the Pacific: 1955*, 1959, *Oceanic Observations of the Pacific: 1956* (The EQUAPAC data), in press; *Oceanic Observations of the Pacific. 1956, 1957, 1958* (The Chinook, Mukluk and Downwind data), in press; H. U. Sverdrup, M. W. Johnson, and R. H. Fleming, *The Oceans*, 1942.

Packaging of equipment

That part of the design process which starts with a functional description of a component or equipment, or "breadboard" developmental assembly thereof, and follows through to the fulfillment of the component or equipment in a form satisfactory for domestic, industrial, or military use. For a discussion of the complete design process *see* SYSTEMS ENGINEERING.

The packaging process includes determining the physical relationships of the elements in the component or equipment, and providing for their mechanical support by chassis, panels, mounts, or other means. In many cases of electronic apparatus, particularly when low si[illegible] and hi[illegible] m[illegible]

plification are present, the arrangement of components relative to each other can strongly influence the internal noise generated in the circuit; packaging decisions must minimize such mutual interference (see CIRCUIT, ELECTRONIC). Thermodynamic and heat transfer considerations require a relationship of the elements and their environment such that the heat resulting from unavoidable inefficiencies in the elements will be adequately dissipated from the component or equipment so as to avoid damage to the assembly (see HEAT TRANSFER). The arrangement, support, isolation, insulation, venting, and sealing of the assembly must be such that its performance satisfies environmental specifications on temperature, humidity, pressure, shock, vibration, acceleration, noise, and other conditions. See ENVIRONMENTAL TEST.

The final physical form of the packaged apparatus varies widely. The package's shape, size, and ht depend on the specifications of the system hich it comprises a part. In some cases a high mum is placed on minimum weight and size, the shape of the package is dictated by seh restricted and sometimes contorted volume tations of the encompassing vessel, as in the of missiles (see MINIATURIZATION OF EQUIPMENT). In other cases, access and maintainability vital considerations; for example, in stationary r transmitters or telephone central office tchboards, units are packaged in more conventional panel form for mounting in vertical racks e MAINTAINABILITY OF EQUIPMENT). The manufacturing techniques to be employed will also demine to a considerable extent the packaged m. See PROTOTYPE (EQUIPMENT).

The reliable service of the packaged equipment ll depend greatly on the talent exercised in the ckaging operation. For a discussion of a number considerations that affect the reliability of comnents and equipments, see RELIABILITY OF EQUIPMENT. [R.W.M.]

acking

seal usually used for high pressure as in steam d hydraulic applications. The motion between arts may be infrequent as in valve stems, or continual as in pump or engine piston rods. There is o sharp dividing line between seals and packing; oth are dynamic pressure resistors under motion. Such diverse materials are used for packing as mpregnated fiber, rubber, cork, or asbestos compounds. The form of the packing may be square, in ring or spiral form, trapezoidal, or V, U, or O-ring in section. In packings, it is necessary that the surface finish of the contacting metal part be smooth for long life of the material. See SEAL, PRESSURE. [P.H.B.]

Packing fraction

The packing fraction of an atom is defined as $f = (M - A)/A$, where M is the mass of the atom in atomic mass units and A is its mass number. The manner in which the packing fraction varies with increasing mass number provides a useful over-all picture of nuclear stability. In any particular mass region, an algebraically smaller packing fraction is an indication of greater stability. The term packing fraction has been largely superseded by the related quantity, binding energy per nucleon. See BINDING ENERGY, NUCLEAR. [H.E.D.]

Packing house

A type of food processing plant generally requiring the use of refrigeration. A meat packing house is a plant engaged in the slaughtering, dressing, and processing of food animals; it uses refrigeration for cold storage and often for freezer storage. A fruit packing house processes fresh fruits, which are usually stored in refrigerated rooms awaiting favorable market conditions. A *precooling plant* is a type of fruit or vegetable packing house which is equipped with refrigeration equipment for rapidly removing the field heat as soon as possible after the product is harvested. See COLD STORAGE; FOOD PRESERVATION; REFRIGERATION. [H.M.HF.]

Paedogenesis

Reproduction by larval or immature animals, especially the parthenogenetic reproduction by larvae of certain gall midges. *Miastor* is a typical example in which this phenomenon occurs. See NEOTENY. [T.I.S.]

Pain, cutaneous

The designation of a group of patterns of somesthetic sensation, generally unpleasant in feeling tone and typically leading to aversive action. Sensations of pain may arise from nearly any part of the skin (cutaneous pain) and from many components of the viscera, muscles, and other deep-lying tissues (deep pain). See PAIN, DEEP.

Specific nature of pain. Because pain may be evoked by intense stimulation of any sense organ, such as dazzling lights, excessively strong tones and noises, pungent odors, and extremes of heat and cold, the error has frequently been made of supposing that pain is simply the result of sensory overstimulation. On the contrary, the evidence is by now overwhelming that sensations of pain owe their existence to a separate system of sense organs and nerves. Just as there is a sense of pressure, one of warm, and one of cold, there is a pain sense; the misleading circumstance is that receptors for pain are almost ubiquitously distributed throughout the body (see CUTANEOUS SENSATION, TEMPERATURE SENSES, TOUCH). The pain of dazzling lights comes, not from the visual system, but from pain fibers terminating in ocular muscles; shrill tones create pain by way of pain endings in the external and middle ear cavities, not through overstimulation of the auditory nerve.

The normal, or adequate, stimulus for cutaneous pain is difficult to specify, the reason being that all classes of stimuli capable of arousing the skin, that is, mechanical, thermal, electrical, and chemical, may evoke pain under some circumstances. It is not a matter of tissue damage, as is often stated, for electrical stimuli too weak to produce injury

reach the pain threshold at lower current values than those needed to elicit pressure sensations. Thermal pain may also be produced without obvious damage to the skin. When such mechanical stimuli as needles, thistle spines, and glass threads are gently pressed into the skin, pain threshold is reached long before penetration of the resistant corneum of the skin can possibly occur. Since the superficial cutaneous layers are pliable, pointed objects create sharp declivities in it. Presumably, the stretching of nerve endings thus produced constitutes the adequate stimulus to pain.

Pain receptors and pain distribution. The only type of cutaneous nerve termination sufficiently distributed throughout the body to serve as a receptor for pain is the free or unencapsulated nerve ending. Nearly everywhere, in the corium and even penetrating the lower reaches of the epidermis, there are free terminations of sensory nerves in the form of skeins, fine twigs, loops, brushlike endings and knobs These frequently overlap and interdigitate with similar structures from nearby fibers. The distribution of pain sensitivity is similarly widespread. For the majority of skin areas, pain points occur in some profusion These are responsive loci which can be found when a spine-tipped horsehair is systematically set down within a circumscribed cutaneous region In one such careful exploration, the density of points varied as follows, measured in pain points per cm²: 232, back of knee; 224, bend of elbow; 203, underside of forearm; 188, back of hand; 144, scalp; 60, ball of thumb; 48, sole of foot; 44, tip of nose. The hollows of the body conformation, the fossae, are especially responsive. Pain sensitivity tends to decrease towards the extremities of limbs, the reverse of the trend governing pressure sensitivity. The two most exquisitely pain-sensitive areas of the body are the cornea of the eye and the inner reaches of the external auditory canal.

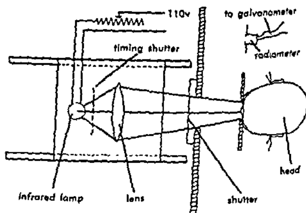
Pain thresholds. There are serious obstacles to the measurement of "pure" pain. If mechanical stimuli are used, they inevitably arouse pressure sensations with a lower expenditure of force than that needed to reach pain threshold. If thermal energy is employed, there are accompanying temperature sensations. Chemical stimuli are difficult to confine to a given skin area. Also, the oily skin surface is highly resistant to penetration by noninjurious reagents. Electrical stimulation, both ac and dc, will arouse pain, but it also spreads and tends to evoke most other cutaneous qualities as well. Algesimeters, instruments for the measurement of pain sensitivity, therefore have to be designed in such a way as to differentiate the pain contribution from other components.

A device currently much in use for experimental and clinical determinations of pain thresholds is a lens focusing radiant heat, and light, on a predetermined spot. A controllable heat source is supplied by an infrared lamp, timing is accomplished by a shutter, and radiation from the lamp is focused on the subject's forehead by means of a condensing lens. In the standard use of the technique, successive exposures are held constant at 3 sec each, w the emitted energy of the lamp being systematica increased from below threshold to the point which the subject feels a single, sharp stab of p just before the closing of the shutter. Warmth is, course, felt throughout much of the exposure riod, and the pain is superposed on this ba ground. By substituting a radiometer in the positi of the forehead the instrument may be calibrat in energy units, that is, in gram-calories per s per cm² of area.

Measurement by this method of the pain thres olds of 150 people, of both sexes and of wide a range, shows an average in the neighborhood 0.21 g-cal/sec/cm² with a variation of about 15 Between this point and 0.48 g-cal/sec/cm² (4 millicalories), there are about 21 discriminab steps of pain intensity. Above 480 millicalories th heat absorbed by the forehead in 3 sec comes clo to burning the skin The threshold for warmth i the same apparatus is far down the intensity scal at about 0.1 millicalorie

Pain adaptation. Pain sensations display th phenomenon of adaptation, in accordance wi which lasting, unvarying stimulation results in dim inution or complete subsidence of the pain. This not readily apparent because of the manner i which headaches, toothaches, and pains from in juries persist until "something is done for them." However, such pains provide poor test cases, since they do not meet the requirement that the stimulation be absolutely unvarying. Most highly disagreeable pains "pound," with rhythmic circulatory changes.

If care is taken to provide a painful stimulus that does not change its intensitive relations to the skin throughout the period of stimulation, the reality of pain adaptation becomes evident. Successful experiments meeting this requirement have been performed with mechanical stimuli (sharp needles), cold stimuli (dry ice), and warm stimuli (radiant heat). In each instance pain diminishes and ultimately disappears, leaving as a residue pressure, cold, or warmth, as the stimulus may dictate.



Radiant heat algesiometer (H G. Wolff and S. Wolf, Pain, Charles C Thomas, 1948)

Quality of pain. If pain is produced in a leg or arm by a pinprick or an intense heat there may be two phases of the resulting feelings. These are "slow" and "fast" pain or "first" and "second" pain. There is elicited immediately "pricking pain," a sharp, suddenly appearing sensation, followed by a slowly rising, much longer lasting "burning pain" The two, easily discriminable from each other, both on the basis of relative delay in appearance and in the pricking burning distinction, have quite different patterns, and must either arise separately or be conducted to the central nervous system in different ways

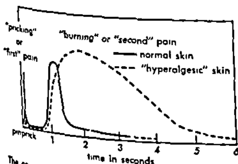
Since the presumption is that the two qualitatively different pains must be set off practically simultaneously at the receptor level, there must be two radically different conduction rates involved This interpretation is confirmed by three sets of facts: (1) If the stimulus is applied somewhere other than at an extremity, such as in the middle trunk region, the time difference evaporates and the two qualities fuse (2) If the conducting nervous pathway is progressively anesthetized with cocaine, the slow, burning pain disappears long before pricking pain does This drug is known to inhibit conduction first in fine, unmyelinated fibers in the array of different fiber sizes of a mixed nerve (3) If a tourniquet is applied to a limb, thus quickly inducing oxygen deprivation in the conducting pathways, pricking pain [illegible]
or dull [illegible]
diamet[illegible]
smaller ones.

The normal difference between pricking and burning pain may be accentuated in instances where severe sunburn or other skin injury renders the tissues hyperalgesic, that is, more sensitive to pain than normal The quick element, first pain, shows no appreciable alteration, but the more slowly developing second pain is both heightened and abnormally prolonged

The *current* conjecture is that slow pain is carried over so-called "C" fibers, having a diameter of less than 1 μ. Fast pain, however, is conducted by fibers of the delta subgroup of the "A" group, having diameters ranging from 3 to 6 μ. Since rate of impulse conduction is known to depend directly on fiber *diameter*, the characteristic delay in the arousal of burning pain is satisfactorily explained.

Pain sensation versus pain reaction. Direct *perception of pain, whether* at threshold level or above, involves only one major reaction of the organism to painful stimuli, the observational response. There are other reactions, some of which tend to interfere with the maintenance of an observational attitude toward pain. Pain is generally regarded as unpleasant, and aversive reactions occur which are designed to reduce or eliminate the source of pain. While cutaneous pain thresholds normally fall within somewhat narrow limits, as seen above, pain reaction, as it is commonly called, is highly variable from one individual to another. As a result of training and of cultural forces, there is a wide range of reaction. For example, there are individuals at one end of the scale who typically "cry before they are hurt" and others, at the opposite extreme, who characteristically withstand even severe pain with the greatest stoicism.

The influences that can reduce or augment the threshold of pain reaction, as opposed to pain perception, are many. These include hypnosis and waking suggestion, an array of analgesic drugs, the subject's mood or attitude, and group mores making for self-chastisement or self-indulgence. Moreover, there are involuntary, visceral responses of the organism that participate in the over-all pain reaction, and some of these can be shown to vary considerably from person to person and also to vary in time in the same individual An example is the galvanic skin response, which is a sudden lowering of the resistance of the skin to an (unfelt) direct electrical current passed through it. This reaction has been studied extensively [F.A C.]



The course of "double pain." (N. Bigelow, I. Harrison, H. Goodell, and H G. Wolff, J Clin. Invest, 1945)

Bibliography: J. D Hardy, H. G. Wolff, and H Goodell, *Pain Sensations and Reactions*, 1953.

Pain, deep

Patterns of somesthetic sensation, originating in the organs of the viscera, muscles, and other deep tissues, generally unpleasant and typically of an aching quality. Whereas cutaneous pain is ordinarily sharply localized, deep pain is usually indefinitely localized and tends to spread as it intensifies (*see* PAIN, CUTANEOUS).

Receptors for deep pain. As in the case of the skin, the only type of nerve termination sufficiently generally distributed throughout the deep tissues to serve as a receptor for pain is the free nerve ending. Pain must necessarily originate in these, though whether the free ending is the only receptor that can perform this function is not known.

The organs of the body cavity, the viscera, are extraordinarily insensitive when judged by criteria applicable to the skin. Surgeons have long known that the abdominal visceral organs may be pinched, squeezed, torn, and cauterized without benefit of anesthetic It has frequently been concluded, therefore, that the viscera are insensitive They are relatively insensitive to such manipulations, to be sure, but their construction is such

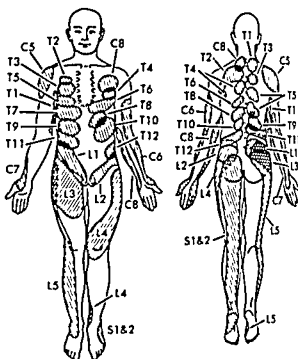
that effective sensory stimulation comes about only though massive tissue distention or severe contraction of musculature. These are the conditions obtaining, for example, in gas pains and cramps. Chemical alteration of muscle bundles, especially through the metabolic imbalance created by circulatory failure, results in severe pain A limb muscle can be excruciatingly painful if exercised while its blood supply is cut off by a tourniquet.

Other deep tissues that have pain endings and which, with sufficiently extensive and prolonged stimulation, can give rise to pain are the major blood vessels, the periosteum (bone covering), mesentery of the intestines, endings of the intercostal and phrenic nerves in the chest and diaphragm, and the mucosal linings of parts of the urogenital system.

Referred pain. Not uncommonly a deep pain originating in a particular organ is felt at some location other than the site of stimulation. For example, in angina pectoris, a pain arises in the heart and would therefore be expected to localize under the sternum, but it may be felt as mainly coming from the chest wall and radiating into the shoulder and underside of the arm. In renal colic a kidney stone passing down the ureter may create an intense pain which is felt not as moving but as being produced steadily in the groin. This localization is the more remarkable when it is considered that the upper end of the ureter actually lies approximately under the last rib

The mechanism of referred pain is understandable only in terms of the segmental mode of organization found in the central nervous system, the spinal cord in particular (*see* SOMESTHESIS). Even though referred pains seem to involve localizations remote from the point of stimulation, it will be found, on tracing the relevant nervous pathways, that it is always the fibers running in the same or neighboring spinal roots that are responsible. Such spread of excitation at the cord level calls in fibers leading to other deep tissues, to superficial skin areas, or to both. Spread to the corresponding region on the opposite side of the body occurs especially easily; pain from a contralateral area is thus a not uncommon symptom when stimulation is intense. Though spread of pain may occur in either direction in the cord, up or down, the possibilities in involvement of collateral fibers are greater in the upward direction.

If the mechanism of referred pain is as described, it should be possible to reproduce artificially some of the faulty localizations by direct stimulation of the spinal roots. This has been done successfully by injecting a 5% saline solution into the appropriate interspinous ligament. This establishes a temporary strong irritation of the spinal center and creates pain in the dermatome belonging to that segment. Injection of the first thoracic segment, for example, induces pain in the shoulder and arm in mimicry of the pains of angina pectoris. Other frequently recurring pains may be simulated in the same way.



Segmental organization of deep pain The numbers identify the region to which pain is referred from injecting hypertonic saline into the cervical (C), thoracic (T), lumbar (L), and sacral (S) interspinous ligaments. (*T Lewis, Pain, Macmillan, 1942*)

Another way in which deep pains can become complicated, especially with respect to localization, is through reflex motor reactions generated by pain impulses. A very general response, one which in turn may provoke pain through biochemical action, is that of blood vessels A vascular spasm may produce anoxia (a state of oxygen need), itself a precipitant of pain, as was seen above.

Hyperalgesia. Sometimes the changed state of excitability created through spinal spread expresses itself not as referred pain but as a generally increased sensitivity of tissues remote from the site of stimulation This so-called' secondary hyperalgesia involves a heightened tenderness of the affected region, one which typically leaves the pain threshold to mechanical or thermal stimuli unaltered but which asserts itself through increased sensation intensity once the pain is aroused.

Primary hyperalgesia also occurs, though it is unrelated to the phenomenon of spread of pain. This is simply the designation for increased sensitivity in the region of pain stimulation itself. The greatly lowered threshold to radiant heat consequent to sunburn or other inflammatory injury is an example. [F A G]

Paint

A fluid, or semifluid, material which may be applied to surfaces in relatively thin layers and which changes to a solid coating with time The change to a solid may or may not be reversible, and may occur by evaporation of solvent, by chemical reaction, or by a combination he two. Paints usu-

ally consist of a vehicle or binder, a pigment which contributes opacity, color, hardness, and bulk to the film, and a solvent or thinner which controls the consistency

Paints may be classified according to the solvent which is, or may be, used for thinning, as solvent-thinned or water thinned; and according to special uses.

Solvent thinned paints. These paints may be classified according to whether the drying mechanism is predominantly solvent evaporation, oxidation or some other chemical reaction Solvent-thinned paints, which dry essentially by solvent evaporation, rely on a fairly hard resin as the vehicle Resins used incl[illegible]

[illegible]

nol, and is commonly used as shellac varnish, without pigment Coat[illegible]

[illegible]

ins usually require a solvent such as a ketone, and even then, only a relatively small amount of material can be dissolved These coatings are usually restricted to those uses in which their unusual properties of color retention and chemical resistance are essential Bitumens or asphalts of petroleum and coal tar origin are often added as fillers to control the flow characteristics when no pigment is needed for opacity. Aluminum or colored pigments may also be used Bituminous coatings are relatively low in cost and are often used for roofing, waterproofing, and the protection of underground pipe lines, where extremely heavy coatings are required, since opportunity for renewal may be limited However, their resistance to weather and to organic solvents is not very high, and they are nearly always dark in color, which limits their fields of application

In paints which dry by oxidation, the vehicle is usually an oil or an oil based varnish These usually contain driers to accelerate the drying of the oil. Paints based essentially on linseed oil, with [su]itable pigments, such as titanium dioxide ex[te]nders, and usually zinc oxide and white lead, are [th]e conventional outside house paints because these [m]aterials give the combination of properties which [me]ets this requirement Where a harder (and usua]lly more brittle) finish is required, as for trim and [in]terior use, an oleoresinous varnish, made from a [d]rying oil and a phenolic or modified phenolic [re]sin, or else an alkyd varnish, made from a dry[in]g oil, glycerin, and phthalic anhydride, is com[m]only used When these materials are finely [g]round, formulated to give a high gloss and good [f]low, and designed for application to smooth sur[f]aces, they are commonly called enamels Flat wall paints are varnish or alkyd based materials, with larger amounts of pigment, for increased wear-re[si]stance, hiding power, and low gloss Primers, [se]alers, and undercoaters are paints that are formu[la]ted to have good adhesion to the substrate and to [fu]rnish a good base for future coats of paint. In many cases they have relatively low amounts of pigment, so that the vehicle will act to seal the underlying surface. Primers for metal usually contain an anticorrosive pigment, such as red lead or zinc chromate

Other paints, which dry essentially by oxidation, include those in which the vehicle is an epoxy ester. The epoxy ester is similar to an alkyd except that the glyceryl phthalate is replaced by an epoxy resin Paints based on modified drying oils, such as those copolymerized with styrene or vinyltoluene, are also included

There are a number of other chemical reactions which may be utilized to produce films. These are usually difficult to produce and control in the exact way desired Therefore, the use of these coatings is restricted to specialized developments. Most of these reactions involve polymerization in one form or another. Urea and melamine resins polymerize by heat and are used in baking finishes where extreme hardness, chemical resistance, and color retention are required, as on kitchen appliances Certain phenolic resins also are converted by heat to yield coatings of excellent water- and chemical resistance

Epoxy, urethane, and polyester resins may be converted, either at room temperature or by baking, but a suitable catalyst is needed The use of these resins requires very close control, but results may be obtained which cannot be reached with other coatings

Water-thinned paints. This group of paints may be subdivided into those in which the vehicle is dissolved in the water and those in which it is dispersed in emulsion form Paints with water-soluble vehicles include the calcimines, in which the vehicle is glue, and casein paints, in which the vehicle is casein or soybean protein Other proteins, including egg albumen, have been used from time to time Because these vehicles are water soluble, the paints are very water-sensitive and are not suitable for exterior use, nor are they washable. Modified oils or resins, which are water-soluble have been introduced These convert to an insoluble form after the evaporation of the water, and are not subject to the above defects

Nearly any solvent-thinned paint may be emulsified by the addition of a suitable emulsifier and adequate agitation However, because the emulsifier remains in the paint film after the water evaporates, these films are usually water-sensitive and have had very little use The development of fugitive emulsifiers and of vehicles especially processed for use in emulsion paints has greatly expanded the use of these materials, because coatings of excellent water-resistance may be obtained and vehicles which have poor solubility in organic solvents may still be suitable for application When the emulsion is formed by emulsion polymerization, the vehicle is usually described as a latex, and these products are called latex paints The most common latexes are made from a copolymer of butadiene and styrene, from polyvinyl acetate, and from acrylic res-

ins. Although there are certain differences among these three materials, products made from them are very similar and they may be treated as a group. Immediately following World War II, latex paints were introduced as flat wall paints because the first ones made had a relatively low gloss and were somewhat water-sensitive. Their ease of application, low odor, alkali-resistance, and the fact that, because they were water-thinned they could be applied to damp plaster or concrete, made them very popular for the decoration of interior walls.

Later, as improvements in formulation were made, these products expanded their field to exterior use, first on masonry and later on wood. Most latex paints are relatively porous and allow moisture trapped below the film to come out without blistering or peeling the film. On the other hand, the high polymers from which latex paints are made do not, ordinarily, have very high adhesion, and some troubles are occasionally experienced when the condition of the substrate is not good. More recent developments include gloss latex paints for either wood or metal, and primers based on latex which are suitable for finishing industrial products, particularly because of the absence of flammable solvents. There are, however, certain difficulties which must be solved before latex paints are widely adopted for these uses. With the development of more satisfactory fugitive emulsifiers, certain other materials, such as alkyds, are also being used as paint vehicles.

Special-use paints. Paints may also be classified by their use. At times only one product is used for a certain purpose, but in general, several different types of materials may be used with excellent results. Architectural paints are usually oil- or varnish-based or latex. Outside house paints are almost always based on linseed oil, although recently, certain alkyd and latex products have been introduced. Exterior trim enamels are either varnish or alkyd. Interior finishes are usually varnish, alkyd, or latex; shellac is widely used for floors. For exterior masonry, a large number of materials may be used, although latex paints are taking over much of the field. Also used is a paint based on portland cement, which dries by reaction with the water, making it very useful for damp areas. Structural metal is usually given one or more coats of an anticorrosive primer, and then topcoats chosen according to the environment. Where the metal is inaccessible, and when appearance is not important, bituminous coatings are frequently used, since very thick films may be applied.

Highway paints. A few examples show the types of products and the factors to be considered for highway paints. The marking of the center lines on highways and other painted areas for the control of traffic requires a paint which dries rapidly in order to avoid interrupting traffic; which adheres well to both asphalt and concrete; which resists abrasion, both wet and dry; and which does not stain from oils, asphalt, or other materials. On the other hand, extreme exterior durability and high flexibility are not required for this service. Solvent thinned paints based on a wide variety of alkyds, modified rubbers, and other resins are used for this work. Adequate traffic paints on latex vehicles have probably not been developed because of relatively poor adhesion to the roadway.

Luminous paints. Luminous paints incorporate pigments which fluoresce under the influence of ultraviolet light. If the luminosity is retained for a period after the exciting light is cut off, these are called phosphorescent paints. Although the intensity is not high, these products are used in decorative schemes and to locate switches and other objects in case of power failure at night. Because nearly all luminous pigments are water-sensitive, a vehicle with extreme water-resistance and impermeability is necessary. For the exciting radiation to be effective, it is also necessary that the vehicle be transparent to ultraviolet light.

Heat-resistant paints. Paints formulated for high heat-resistance are another specialized form. Although some conventional materials will resist temperatures up to about 300°F for a period of time, for resistance to higher temperatures, different approaches are required. One method used is to incorporate an aluminum pigment in a vehicle which will burn off in time. In this manner, the aluminum is fused to the surface, and a certain amount of protection is obtained. If good weather-resistance is also required, a silicone vehicle is usually used. The pigments must also be chosen for their resistance to heat in the temperature range required.

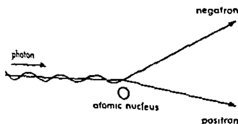
Corrosion-resistant paints. For the protection of chemical plants, and in other areas where relatively high concentrations of chemicals, solvents, and corrosive fumes may be encountered, there are a number of specialized coatings available. These include coatings based on rubber and rubberlike materials, epoxy coatings, and coal-tar derivatives.

Factory-applied finishes. In the field of factory-applied finishes, the same specialized consideration must be given to each problem with the additional difficulty that the product must be tailored to the particular production line in which it is to be used. In the finishing of furniture, lacquers are usually suitable, sometimes varnishes and shellac. Epoxy and unsaturated polyester coatings have recently been adopted in certain fields. Automobile manufacturers have used lacquers and alkyds until the recent adoption of acrylic and urea coatings. Urea coatings which approximate the properties of vitreous enamels are widely used for kitchen and bathroom appliances. Almost the entire range of products of the industry, in some form or combination appears in these factory-applied finishes. *See* DRIER (PAINT); DRYING OIL; PIGMENT; POLYMER, SOLVENT; SURFACE COATING; THINNER; VARNISH. [F.SO.]

Bibliography: W. Von Fischer (ed.), *Paint and Varnish Technology*, 1948.

r production (electron-positron)

rocess in which a negative electron (negatron) a positive electron (positron) are simultane-ly created in the vicinity of a nucleus or an ele-itary particle. In external pair production, an tromagnetic wave (photon) is absorbed and ites an electron pair. The absorption of high-rgy γ rays is due mainly to this effect (see gram). Internal pair production is not associ-d with observable electromagnetic radiation and y occur when an excited nucleus releases some ts internal energy.



ernal pair (negatron-positron) production.

Pair production is of considerable theoretical erest, not only as an example of the materializa-n of energy, but also as a striking confirmation the relativistic quantum theory proposed by A M Dirac This theory has made possible antitative predictions of production probability, ferential electron distribution, and kinetic en-gy partition. The results are in satisfactory agree-ent with experimental findings See QUANTUM EORY, RELATIVISTIC.

External pair production can take place only if e energy of the photon exceeds $2mc^2$ (m = elec-on mass, c = velocity of light) or 1 02 Mev, hich is the energy required for production of an lectron pair at rest The energy excess, $h\nu - 2mc^2$ ν is the frequency of the photon, h is Planck's onstant), appears as kinetic energy of the created articles The sharing of this energy between the lectrons takes place in a random way, such that he positron, for example, may assume any energy etween zero and $h\nu - 2mc^2$ with about the same robability Because of the electrostatic repulsion of the nucleus, the positron actually obtains a igher energy on the average than the negatron

Conservation laws require that the momentum of the initial photon be transferred to the product particles (For a discussion of conservation laws, see NUCLEAR REACTION.) Simple calculations show that this can be fulfilled only if a third particle or system of particles takes part in the process. It may be a nucleus, as is usually the case, but in principle any charged particle may restore the momentum balance. For a given energy division between the electrons, the nucleus may recoil in any direction; consequently, the direction in which the electrons are emitted is not fixed, but is randomly distributed As a consequence of its large mass, the nucleus receives only a vanishingly small part of the initial photon energy.

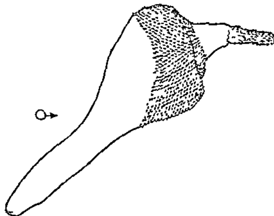
Internal pairs are often emitted from radioactive substances. After radioactive decay, the daughter nucleus may be left with excess energy. Although this energy is usually released as electromagnetic radiation, pair production may compete when the energy available exceeds $2mc^2$, the probability being higher the higher the energy release. The angular correlation of the pairs and the production probability also depend on the multipole order of the transition. See MULTIPOLE RADIATION; see also GAMMA RAYS; POSITRON, QUANTUM FIELD THEORY. [C.B.]

Bibliography: R. D Evans, *The Atomic Nucleus*, 1955.

Palaeacanthocephala

An order of the Acanthocephala, the adults of which are parasitic worms found in fishes, aquatic birds, and mammals They have the following characteristics. The nuclei of the hypodermis are fragmented and the chief lacunar vessels are lateral. The males have less than eight but more than one cement gland The ligament sac in the female breaks down so that the eggs develop in the body cavity. Proboscis hooks occur in long rows and spines are present on the body of some species. Species which commonly occur in vertebrates are *Leptorhynchoides thecatus* and *Corynosoma*.

Leptorhynchoides thecatus. This is one of the most common species of acanthocephalan in North American fresh-water fish The body is a creamy color, long, slender, and devoid of spines. Both ends of the body are curved ventrally. The females are 6–26 mm and males 3–12 mm in length. Females usually are thicker than the males. The proboscis is long, slightly enlarged in the middle, and bent ventrally at an angle to the body Proboscis hooks are quincuncial in arrangement with 12 longitudinal rows of 12–16 hooks each. The base of the hook is enveloped by a prominent cuticular



Corynosoma reductum (From H J Van Cleave, *Acanthocephala of North American Mammals*, Univ of Illinois Press, 1953)

sheath. Genital organs of the male occur in the posterior half of the body. Eight cement glands usually are arranged in three levels. The eggs are spindle-shaped with the middle membrane or shell thicker because of polar enlargements. The outer membrane is thick and refractive. The intermediate host is an amphipod, *Hyalella azteca*. Adults have been recorded from the ceca and small intestine of 79 species of fish.

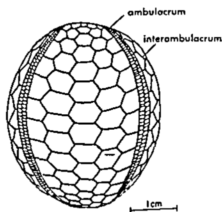
Corynosoma spp. Individuals of the genus *Corynosoma* reach sexual maturity in the small intestine of birds and mammals with aquatic habitats. The body is club-shaped and 2–10 mm long; the anterior end is thickened as an inflated bulb, whereas the hind trunk is narrow and cylindrical. The body is provided with spines on the anterior extremity of the trunk, which extend farther along the ventral surface than the dorsal. In some species the trunk spines extend the full length of the trunk on the ventral surface. The body spines are of various forms, often sigmoidal, and the tip of each is commonly invested by a cuticular fold. The proboscis is directed ventrally. Proboscis hooks occur in longitudinal rows, and the individual hooks increase in thickness from anterior to posterior. The male organs are restricted to the posterior half of the body with the testes rounded or slightly elongate Six cement glands, pyriform to clavate in shape, are present. Eggs are spindle-shaped with a short axial prolongation of the middle membrane. The species of this genus are distributed through all continents Aquatic mammals (seals) and water birds (ducks) are the normal definitive hosts, although fishes serve as the second intermediate or transport host and various crustaceans as the first intermediate host *See* ACANTHOCEPHALA.

[D V.M.]

Palaechinoida

An extinct order of Perischoechinoidea with a rigid test in which the ambulacra bevel over the adjoining interambulacra There are two or more columns of plates in the ambulacra, three or more in the interambulacra. (An exception is one only in *Cravenechinus.*) The order, which is known only from the Lower Carboniferous, includes such generalized forms as *Palaechinus* (see illustratic and more specialized forms such as *Melonechu* and *Cravenechinus*. *See* PERISCHOECHINOIDEA

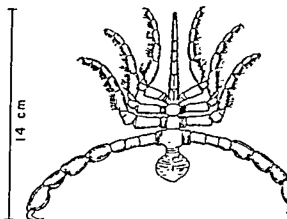
[H B



Palaechinus ellipticus, Lower Carboniferous of Europe.

Palaeoisopus

A peculiar arthropod of uncertain position, rej sented by several well preserved fossils from Devonian Hunsruck shales. Considered by some



Palaeoisopus problematicus. (After Broili)

be a fossil pycnogonid, its general structure, w jointed anterior process, bulbous posterior ab men, and posterior paddlelike legs, makes it possible to classify it with any Recent arthroj group. *See* PYCNOGONIDA.

[J W

Bibliography: J. W. Hedgpeth, *Palaeoisopus*, R C. Moore (ed), *Treatise on Invertebrate Pale tology*, pt P, 1955

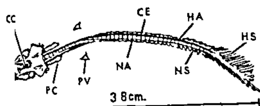
Palaeonemertini

An order of the class Anopla in the phyl Rhynchocoela Most of the palaeonemertines littoral species Three families are recogniz Tubulanidae, *containing the genera Tubulan Carinina*, and *Procaranina*; Carinomidae, which monogeneric, *Carinoma*; and Cephalothridid with the main genera *Cephalothrix* and *Procepl lothrix* This order is separated from the hete nemertines on the basis of the body wall muscu ture which may be either two- or three-layer Both a middorsal vessel and connectives betwe the lateral vessels are usually lacking, as are oce and cerebral organs. *See* ANOPLA, HETERONE ERTINI.

[C B

Palaeospondyloidea

An ordinal name assigned to the single, tiny, prc lematic fish *Palaeospondylus* which is known on from Middle Devonian shales of restricted exte in Caithness, Scotland This animal seldom e ceeded 5 cm in length. The skeleton consists of well-calcified skull, vertebral column, caudal fi and occasional traces of paired pelvic fins The r lationships of *Palaeospondylus* are uncertain has been considered variously to be an agnathoi fish, an elasmobranch, a al placoderm, a larv



keleton of *Palaeospondylus*, a small and problemati- al Middle Devonian fish, showing dorsally· CC, roughlike cranial cavity; PC, presumed pectoral gir- le, PV, pelvic girdle, NA, neural arch; CE, centrum, IA, hemal arch, NS, neural spine; HS, hemal spine After J A Moy-Thomas)

lungfish, and a larval amphibian. For the present this fish is most widely regarded as a placoderm with completely undeveloped dermal armor. *See* PLACODERMI. [D.H.D.]

Palaeotremata

An order of the brachiopod class Articulata, in which two superfamilies, Rustellacea and Kutorginacea, are included The Rustellacea gave rise to the Kutorginacea The round to transversely rounded, thick shells of these brachiopods are chitinous and calcareophosphatic in the superfamily Rustellacea and are calcareous in the Kutorginacea. They are primitive Articulata, without



Kutorgina cingulata (Lower Cambrian, Vermont), side, ventral, and dorsal views, actual size (From W H Twenhofel and R R Shrock, *Principles of Invertebrate Paleontology*, McGraw-Hill, 1953)

fully developed hinge teeth and corresponding sockets, or delthyria The pedicle protrudes from between the valves The exterior exhibits concentric growth lines and the valves are convex Muscle scars indicate that these forms are ancestral to the Protremata. Geologically, they range from the Cambrian to the Ordovician. *See* ARTICULATA (BRACHIOPODA) [K H.]

Palate

The roof of the mouth in those vertebrates whose mouth cavity and nasal passage are kept wholly or partially separate The nasal cavities of fishes are independent, blind sacs In amphibians, on the other hand, these nasal cavities communicate with the front of the mouth by a pair of apertures, the internal nares Such an arrangement is adapted to the breathing of air when the mouth is closed Reptiles and birds possess a pair of bony flanges, separated by a median cleft so that the incomplete

... for the free ... nds from ... back to the pharynx. Crocodiles and mammals create a wholly separate channel by interposing a complete palate between the air passages and the mouth cavity. The definitive communication of the air channels with the nasal pharynx is known as the secondary choanae. Teeth usually occur on the roof of the mouth of bony fishes, amphibians, and reptiles. Many mammals, and especially hoofed and carnivorous forms, have the palate set with transverse ... which aid in the ma...

that serve to strain out minute organisms to be eaten.

The hard palate. The palate of mammals consists of two portions. The hard palate, more anterior in position, underlies the nasal cavity, whereas the soft palate hangs like a curtain between the mouth and nasal pharynx The hard palate has an intermediate layer of bone, supplied anteriorly by paired palatine processes of the maxillary bones, and posteriorly by the horizontal part of each palate bone The oral surface of the hard palate is a mucous membrane, covered with a stratified squamous epithelium Anteriorly in man there are four to six transverse palatine ridges; these diminish in prominence between fetal life and old age. A submucosal layer bears pure mucous glands and binds the membrane firmly to the periosteum of the bony component. Above the bone is the mucous membrane that constitutes the floor of the nasal cavity There is a falsely stratified, ciliated epithelium underlaid by mixed seromucous glands. Nearest to the periosteum of the bone is a layer of elastic fibers

The soft palate. The soft palate is a backward continuation from the hard palate Its free margin connects on each side with two folds of mucous membrane, the palatine arches, enclosing a palatine tonsil In the midline the margin extends into a fingerlike projection named the uvula. Both the hard and soft palate bear a seam, or raphe, along the midline. The oral side of the soft palate continues as the covering of the hard palate, and the submucosa contains pure mucous glands. The intermediate layer is a sheet of voluntary muscle, to which several palatal muscles contribute The nasal side continues the structures described for the hard palate but posteriorly, near the free margin, the epithelium becomes stratified because it makes contact at times with the nasopharynx.

Besides separating the nasal passages from the mouth, the hard palate is a firm plate, against which the tongue crushes and manipulates food The soft palate, at rest, is pendent. In sucking, swallowing, or vomiting it is raised to separate the oral from the nasal portion of the pharynx This closure prevents food from passing upward into the nasopharynx and nose. The closing action also

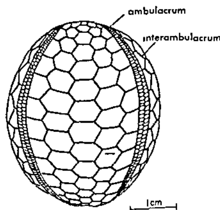
sheath Genital organs of the male occur in the posterior half of the body. Eight cement glands usually are arranged in three levels. The eggs are spindle-shaped with the middle membrane or shell thicker because of polar enlargements. The outer membrane is thick and refractive. The intermediate host is an amphipod, *Hyalella azteca*. Adults have been recorded from the ceca and small intestine of 79 species of fish.

Corynosoma spp. Individuals of the genus *Corynosoma* reach sexual maturity in the small intestine of birds and mammals with aquatic habitats The body is club-shaped and 2–10 mm long; the anterior end is thickened as an inflated bulb, whereas the hind trunk is narrow and cylindrical. The body is provided with spines on the anterior extremity of the trunk, which extend farther along the ventral surface than the dorsal. In some species the trunk spines extend the full length of the trunk on the ventral surface. The body spines are of various forms, often sigmoidal, and the tip of each is commonly invested by a cuticular fold. The proboscis is directed ventrally. Proboscis hooks occur in longitudinal rows, and the individual hooks increase in thickness from anterior to posterior. The male organs are restricted to the posterior half of the body with the testes rounded or slightly elongate. Six cement glands, pyriform to clavate in shape, are present Eggs are spindle-shaped with a short axial prolongation of the middle membrane. The species of this genus are distributed through all continents. Aquatic mammals (seals) and water birds (ducks) are the normal definitive hosts, although fishes serve as the second intermediate or transport host and various crustaceans as the first intermediate host See ACANTHOCEPHALA

[D V.M]

Palaechinoida

An extinct order of Perischoechinoidea with a rigid test in which the ambulacra bevel over the adjoining interambulacra. There are two or more columns of plates in the ambulacra, three or more in the interambulacra. (An exception is one only in *Cravenechinus*.) The order, which is known only from the Lower Carboniferous, includes such generalized forms as *Palaechinus* (see illustration) and more specialized forms such as *Melonechinus* and *Cravenechinus*. See PERISCHOECHINOIDEA.

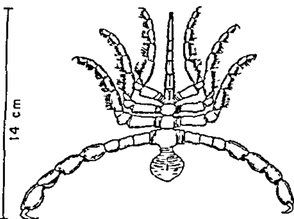


Palaechinus ellipticus, Lower Carboniferous of Europe.

[R B F]

Palaeoisopus

A peculiar arthropod of uncertain position, represented by several well preserved fossils from the Devonian Hunsruck shales Considered by some to be a fossil pycnogonid, its general structure, with jointed anterior process, bulbous posterior abdomen, and posterior paddlelike legs, makes it impossible to classify it with any Recent arthropod group See PYCNOGONIDA



Palaeoisopus problematicus. (After Broili)

[J.W H.]

Bibliography: J W. Hedgpeth, *Palaeoisopus*, in R C. Moore (ed), *Treatise on Invertebrate Paleontology*, pt. P, 1955.

Palaeonemertini

An order of the class Anopla in the phylum Rhynchocoela. Most of the palaeonemertines are littoral species Three families are recognized: Tubulanidae, containing the genera *Tubulanus*, *Carinina*, and *Procaranina*, Carinomidae, which is monogeneric, *Carinoma*, and Cephalothricidae, with the main genera *Cephalothrix* and *Procephalothrix*. This order is separated from the heteronemertines on the basis of the body wall musculature which may be either two- or three-layered. Both a middorsal vessel and connectives between the lateral vessels are usually lacking, as are ocelli and cerebral organs See ANOPLA; HETERONEMERTINI.

[C B C]

Palaeospondyloidea

An ordinal name assigned to the single, tiny, problematic fish *Palaeospondylus* which is known only from Middle Devonian shales of restricted extent in Caithness, Scotland This animal seldom exceeded 5 cm in length The skeleton consists of a well-calcified skull, vertebral column, caudal fin, and occasional traces of paired pelvic fins The relationships of *Palaeospondylus* are uncertain. It has been considered variously to be an agnathous fish, an elasmobranch, a larval placoderm, a larval

of the fossils would have decomposed since the fossil was laid down Carbohydrates, cellulose, fats, and porphyrins also have been found in fossils. See AMINO ACIDS

The organic components of living organisms in the past were surely the same as those in organisms which live today. Proteins, nucleic acids, carbohydrates, and lipids occur in all presently living organisms, from aerobic and anaerobic bacteria to mammals Even the biochemical pathways of intermediary metabolism of all aerobic organisms are very similar. Differences do occur as the evolutionary scale is ascended, but these differences are not numerous and are mostly differences in excretory products and the abilities to synthesize vitamins and amino acids It is very likely that the metabolism of an existing species is almost identical with the metabolism of the same species millions of years ago *See* CARBOHYDRATE; LIPID; METABOLISM, NUCLEIC ACID; PROTEIN.

In view of this it is probable that the fossils when laid down had the same constituents as present organisms. Depending upon the conditions of deposition, ground water might leach out some of the organic compounds, and bacteria might utilize some of the compounds The remaining compounds will be found in the fossil if they have not decomposed since they were laid down

All organic compounds are thermodynamically unstable and will eventually decompose However, the rate of decomposition will vary greatly depending upon the temperature and activation energy for the decomposition By a careful study of the stabilities of biological organic compounds and the compounds found in a fossil, it may eventually be possible to tell something of the thermal history of a fossil of known age by the compounds found in it. *See* PALEOECOLOGY (GEOCHEMICAL ASPECTS); *see also* FOSSIL; SOIL MICROBIOLOGY.

[S L M.]

Paleobotany

The study of fossil plants and vegetation of the geologic past. As a branch of paleontology, it combines a knowledge of both geology and botany. Its materials are the fossilized remains of prehistoric plants preserved in the rocks, including fossil leaves, seeds, spores, and wood, and occasionally fruits and flowers. From such materials the paleobotanist attempts to reconstruct whole plants as they actually grew, as well as the ancient forests of which they were a living part. The fossil plant record shows the succession of plants which have inhabited the earth, from the relatively simple aquatic forms of the older geologic eras to forms progressively more complex and better adapted to land habitats in the younger geologic eras Paleobotany involves not only collecting, describing, and naming fossil plants, but also aims at their interpretation in terms of the evolutionary history of plant groups, the relationships between extinct and living forms, the reconstruction of the environmental conditions under which the ancient forests lived, and the relative geologic ages of the rocks in which the fossils were found.

PLANT FOSSILS

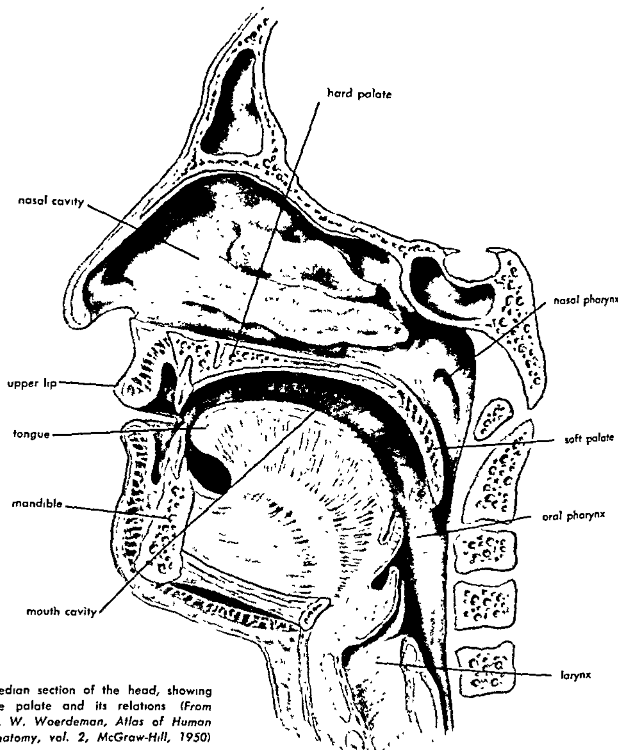
How fossil plants are preserved. In living forests accumulations of plant debris normally disappear as the result of the natural processes of disintegration and decay. However, when forests grow in or adjacent to areas of sedimentary deposition, such as swamps, lakes, streams, and estuaries, plant debris may be buried along with the sediments. In the course of time and under the proper geologic conditions such sediments become sedimentary rocks and the buried plant remains become fossilized. The plant fragments most likely to be preserved as fossils are those possessing hard tissues that resist decay. The most resistant are the wax-coated spores and pollen grains, followed in relative order by nuts and seeds, wood, leaves, and finally flowers. *See* FOSSIL SEEDS AND FRUITS; PALYNOLOGY.

Not all environments of deposition are equally conducive to the preservation of fossil plants. Deposits laid down in upland lakes, streams, or swamps, for example, are easily eroded away in the course of time, whereas lowland deposits are much less apt to be removed; the majority of fossil plants occur in deposits of the latter type. The deposits of the ocean bottoms, on the other hand, rarely contain the remains of land plants This is mainly because of the destructive action of waves and currents along coastlines where plant remains might otherwise enter the sedimentary record. The remains of seaweeds and related aquatics are therefore usually the only kinds of plant fossils found in marine deposits

Sources of plant fossils. Most fossil plants are found in shales and fine-grained siltstones, which are the lithified muds of ancient deposits The best collections are obtained from rocks originally laid down as deposits in flood-plains, lakes, swamps or bogs, and coastal lagoons or estuaries. Volcanic ash, the lithified form of which is known as tuff, is also an excellent medium for the rapid burial and subsequent preservation of plant remains. Only rarely do associated volcanic lavas preserve the casts of tree trunks or the impressions of leaves, due to their destructive high temperature at the time of eruption.

Coal and associated beds. Since coal is known to be altered plant materials originally accumulated in swamps, it is natural that much knowledge of ancient plants is derived from a study of coal itself and of the fossil plants that occur both in coal and in the roof shales and underclays associated with coal beds *See* COAL; COAL BALLS; COAL PALEOBOTANY

Fossilization processes. After burial by sediments, plant remains may undergo any one of several different processes leading to ultimate fossilization: (1) they may remain essentially unaltered or only partially altered, forming compressions; (2) they may be completely removed, leaving only



Median section of the head, showing the palate and its relations (*From M. W. Woerdeman, Atlas of Human Anatomy, vol. 2, McGraw-Hill, 1950*)

occurs in speech, except for certain consonants requiring nasal resonance. The soft palate can also be lowered into contact with the root of the tongue The palate develops from lateral folds of the primitive upper jaw that meet and fuse in the midline Bone differentiates in the front half and muscle in the remainder. When the process of fusion fails to any degree along its course, there results a malformation known as cleft palate. *See* SPEECH.

[L B.A.]

Paleobiochemistry

The biochemistry of organisms which lived in the past, particularly the organisms found in fossil form. The work in this new field has been confined to examination of fossils for organic compounds, and to measuring the stabilities of biological organic compounds which are present or have decomposed in these fossils. The formation of petroleum from organic material is a related topic. *See* PETROLEUM (ORIGIN).

Amino acids have been found in fossils varying in age from several thousand years to 360,000,000 years In fossils over 1,000,000 years of age, only glycine, alanine, valine, leucine, and glutamic and aspartic acids are found; the other amino acids have decomposed. When amino acids are tested for thermal stability, it is found that these amino acids are the more stable ones, and that the other amino acids which almost surely occurred in the protein

ie silica by immersion in hydrofluoric acid leaves soft, spongy mass of woody tissue which can be nbedded and sectioned for study by the same chniques botanists use to study modern wood

The more common minerals producing petrifacons are quartz (silica, SiO_2), calcite (calcium rbonate, $CaCO_3$), pyrite and marcasite (iron lfides), and various iron oxides. *See* PETRIFACON, PETRIFIED FORESTS

Significance of fossil plants. To the geologist, ie record of fossil plants added to that of fossil imals supplies the knowledge of the changing inorama of life in earth history. The study of the

gic ages of the rocks in which the fossils are und. Where the geologic time range of a plant

roups during the eras, *periods*, and epochs of the ologic past are shown in Fig 1 In the study of storical geology the fossil plant record helps to mplete the total picture of the ancient geography paleogeography) of each of the successive chapers of the past.

Fossil plants are also reliable indicators of past ironmental conditions and are often called the ermometers of the past It is a well-known fact at in modern floras certain plants are restricted rather specific climatic and topographic conditions. If fossilized plants can be shown to be closely lated to such living plants of restricted range. it ay be assumed that the fossil plants lived under milar restricted conditions The finding of fossil lms laurels, magnolias, peppers, and cycads in e early Tertiary rocks of southeastern Alaska, r example, is evidence of subtropical to warm mperate, lowland conditions there at that time. ee POSTGLACIAL VEGETATION AND CLIMATE.

To the botanist, fossil plants supply a knowledge f the numerous past developments within the vegeable kingdom which have led to present conditions. he fossil plant record has furnished many missing iapters to the story of the slow, unending evoluon of plants from the simpler forms of earlier ologic ages to the more complex plants of the resent day The recognition that living vegetation merely the end result of major changes in disibution, as well as of evolution in the past, has led to a better understanding of modern plant ography The present restricted distribution of ch forms as the redwoods (*Sequoia*), the dawn dwoods (*Metasequoia*), and the Oriental *Ginkgo* an be properly appreciated only in terms of their despread distribution in the past Still other lants including several large groups such as the ed ferns (Pteridospermae) and the cycadeoids Cycadeoidea), are shown by their fossil record to ave flourished for many millions of years and then eclined to complete extinction, a phenomenon of pecial interest to plant geneticists

The fossil plant record has also been responsible for a better understanding of plant classification (taxonomy). The most widely accepted scheme of classification at present has been based almost as much on a knowledge of fossil plants as of living plants *See* PLANT CLASSIFICATION.

CLASSIFICATION OF FOSSIL PLANTS

Insofar as possible, fossil plants are classified according to the natural classification used by both botanists and paleobotanists. The basis for this classification consists essentially of three morphological characters: (1) *the nature and relationship* of leaf and stem, (2) the anatomy of the stem; (3) the arrangement *and position* of the sporebearing organs The classification presented below is a modern modification of older systems. It reflects the influence of a knowledge of fossil plants on *concepts of classification*. The fossil record shows, for example, that the presence of seeds, previously considered to be an evidence of affinity, has been developed independently in such diverse groups as seed ferns, conifers, cycads, and flowering plants In the following major subdivisions of the plant kingdom the names followed by an asterisk (*) are those groups known only in the fossil state.

Scientific names	*Common names*
Thallophyta	
Algae	Seaweeds and allies
Fungi	Fungi
Bryophyta	
Hepaticae	Liverworts
Musci	Mosses
Tracheophyta	
Psilopsida	
Psilophytales*	Psilophytes
Psilotales	
Lycopsida	
Lepidodendrales*	Scale trees
Pleuromeiales*	
Lycopodiales	Lycopods, club mosses
Sphenopsida	
Hyeniales*	
Sphenophyllales*	Sphenophylls
Calamitales*	Calamites
Equisetales	Horsetail rushes
Pteropsida	
Filicineae	Ferns
Coenopteridales*	Ancient ferns
Filicales	Modern ferns
Gymnospermae	
Cordaitales*	Cordaits
Pteridospermae* (Cycadofilicales)	Seed ferns
Cycadeoidales* (Bennettitales)	Cycadeoids
Cycadales	Cycads
Ginkgoales	Ginkgos
Coniferales	Conifers
Caytoniales*	
Angiospermae	Flowering plants
Monocotyledoneae	Monocots
Dicotyledoneae	Dicots

impressions or molds, and sometimes casts; or (3) they may be partially altered and subsequently permeated by mineral-bearing solutions, producing petrifactions.

Compressions. Most plant fossils occur as compressions formed when leaves, seeds, flowers, or other plant remains are compressed between layers of sediments or their lithified equivalents. Only rarely do plant materials remain essentially unaltered (mummified): examples include specimens of fresh plant debris from frozen muds of polar regions and the asphalt-impregnated leaves, cones, seeds, and wood fragments found in the Rancho La Brea tar pits of California. Also essentially unaltered are spores and pollen grains of microscopic size, whose waxy outer coats enable them to be preserved in countless numbers in many coals and nonmarine shale beds. Slightly more altered are compressions in which only the resistant carbon is left. Such carbonized leaves, woody trunks, and branches are common in the impervious clays of New Jersey despite their age of more than 80,000,000 years. In the still older paper coals of Russia, in deposits 240,000,000 years old, carbonized leaves retain their outside cuticle and may be peeled from their clay medium and mounted in plastic or between two pieces of glass.

Molds, casts, and impressions. More or less complete decay and removal of plant remains from their enclosing rock may leave only an empty cavity or mold. If such molds subsequently fill in with sediments or mineral deposits, casts may be formed producing replicas of the original plant material. Examples of these are common in the vicinity of hot springs where standing trees are killed and entombed by the deposition of travertine. Molds of plant fragments including delicate flowers are also found in amber, which is hardened resin of ancient gum-bearing trees. See AMBER; TRAVERTINE.

In some cases leaves, needles, seeds, or flowers may be completely dissipated and leave only imprints in their enclosing medium. These are known as impressions, which are common in the coarser pervious rocks in which active ground water movement has facilitated decay and removal of the buried plant materials.

Petrifactions. In some instances ground water may carry high concentrations of mineral matter in solution. Thes[illegible] culate throug[illegible] or rocks and [illegible] though it was [illegible] are *fossils in which the original plant material* is replaced, molecule by molecule, by mineral matter this is actually *not* the case. The process of petrifaction is now believed to involve the slow infiltration of solutions through the buried plants and the gradual filling in of the minute cell cavities and intercellular spaces with mineral matter. The original woody components of the cell walls are not replaced but are rather surrounded by mineral matter and remain so nearly intact that they can be recovered by the chemical removal of the petrifying minerals. In the case of silicified wood, removal of

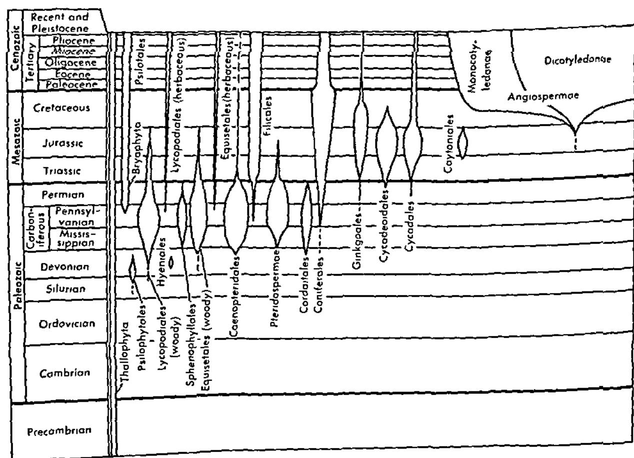


Fig. 1. Geologic distribution of plant groups

their ancestry back to the coal swamps of the late Paleozoic Era The ancient ferns (Coenopteridales), differing mainly in stem structure and arrangement of spore bearing organs, were dominant in the Pennsylvanian forests. Some of the living Ophioglossaceae and Marattiaceae may be descendants of the ancient ferns, most of which became extinct by the end of the Permian Period *See* COENOPTERIDALES; FILICALES.

Gymnospermae These are plants characterized by so-called naked seeds, and include seven orders of trees and shrubs, of which four orders are extinct.

Cordaitales The most ancient order, the Cordaitales are believed ancestral to the later conifers from which they differ mainly in their long straplike leaves and their more open, conelike, seed-bearing organs The cordaits were common among the forest trees of the late Paleozoic Era; they are not known to have survived into the Mesozoic Era. *See* CORDAITALES

Pteridospermae. Another ancient group, the Pteridospermae, or seed ferns, differ from true ferns in the possession of well-defined seeds usually borne at the ends of branches of the fernlike fronds They were among the dominants of the late Paleozoic Era, after which they declined rapidly and became extinct before the end of the Jurassic Period *See* PTERIDOSPERMAE.

Cycadeoidales. The cycadeoids were essentially restricted to the Mesozoic Era They resembled most living cycads in both foliage and internal stem structure but differed greatly in the nature of their flowerlike inflorescences These developed laterally on the trunks among the leaf bases, often as many as 200 flowers on a single plant *See* BENNETTITALES

Cycadales The modern cycads are widespread in the present tropics where they are often called sago palms. In contrast to the cycadeoids, their reproductive structures resemble more closely the cones of living conifers The group has a long geologic history extending as far back as the Permian. Its climax was reached during the middle of the Mesozoic Era. *See* CYCADALES.

Coniferales. The conifers, which today are the most abundant and widespread of the Gymnospermae, are chiefly forest trees with needle-shaped or awl-shaped leaves and with seeds borne in true cones The earliest conifers, belonging to extinct genera, occurred only rarely in the late Paleozoic Era Increasing in abundance and distribution in the Triassic Period, the group attained its greatest development in the Jurassic and Cretaceous Periods (Fig. 1). Since then the conifers have declined, presumably as the result of competition from the flowering plants. *See* CONIFERALES.

Ginkgoales The Ginkgoales, at present restricted to a single species of Oriental *Ginkgo*, was a dominant group during most of the Mesozoic Era *See* GINKGOALES

Caytoniales. The Caytoniales, known only in the fossil state, were a small mid Mesozoic group whose fruits bore a resemblance to both the older seed ferns and the younger flowering plants. *See* CAYTONIALES

Angiospermae. The flowering plants, in addition to having flowers, are characterized by enclosed seeds and by broad, flat leaves like those of palms, oaks, maples, and elms (Fig 3). They are not reliably known before the Late Jurassic Period.

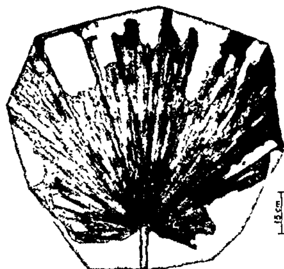


Fig 3 Impression of an angiosperm leaf, a fan palm, from the Eocene of Wyoming

Increasing and spreading slowly during the Early Cretaceous, they increased tremendously during the Late Cretaceous, reaching almost their present dominance before the beginning of the Cenozoic Era. *See* ANGIOSPERMAE

FLORAS OF THE GEOLOGIC ERAS

Precambrian floras. The fossil remains of both plants and animals are extremely rare in rocks of Precambrian age, which are usually referred to an older Archeozoic Era and a younger Proterozoic Era There is no reliable evidence of the existence of land plants during this portion of the earth's history; even marine plants are scarce and are restricted chiefly to the later phases of the Proterozoic Era In the Archeozoic rocks of Finland, a few occurrences of small bits of true carbon are believed to represent the remains of primitive algae (*Corycium*) more than 2,000,000,000 years old. Canadian rocks of approximately the same age have yielded specimens of both filamentous algae and fungal hyphae with attached spores In the younger Proterozoic limestones of Montana and adjacent Canada numerous spherical, dome-shaped, and columnar masses of concentric laminations are believed to be the remains of ancient algal reefs. Similar forms, referred to the genus *Collenia*, are widely known from rocks of the same age in Alaska, Ontario, Greenland, and the Grand Canyon, Arizona Specimens of blue-green algae and bacteria have been described from the Lake Superior iron ores In the Adirondack region, beds of Precambrian graphite are interpreted as metamorphosed

The assignment of fossil plants to their proper groups within the plant kingdom often presents difficulties not encountered with modern plants. For example, the identification of species, genera, and families among the living flowering plants depends in large part on characteristics of the flowers, which occur only very rarely as fossils. In this group, therefore, the identification of fossil forms must rely on those portions which are actually found fossilized, such as leaves, wood, fruits, and spores.

Another difference encountered in the classification of fossil plants and modern plants is a result of the fragmented condition of most fossil plants. Such detached parts, particularly those belonging to extinct plant groups, often cannot be certainly associated or reconstructed into a complete plant. This makes it necessary and convenient to institute so-called form genera for different portions of what may later prove to be a single plant. For example, in Carboniferous rocks segmented, ribbed stems may be referred to the form genus *Calamites*, the foliage to *Annularia*, and the spore-bearing cones to *Calamostachys*.

Form genera of another type may refer to fossil plants whose morphological characters show only a family relationship to modern forms; *Magnoliophyllum*, for example, implies a general resemblance to the magnolia family without a commitment of closer relationship to any particular genus within the family.

GEOLOGIC DISTRIBUTION OF PLANT GROUPS

The fossil record of the plant kingdom shows that the major plant groups have varied considerably in time of origin, geologic range, and period of maximum development and decline (Fig. 1). Clearly demonstrating the theory of evolution, the earliest plants belong to the simpler groups and the later ones become progressively more complex. Modern land vegetation is thus a composite made up chiefly of the most highly complex plants, the flowering plants, living in association with a lesser number of survivors of simpler types of plants.

Thallophyta. This group of plants is composed mainly of the marine seaweeds. They are not abundant in the fossil record because of their soft, perishable nature. Some of the lime-secreting algae, however, are more resistant to destruction and so are fairly abundant, especially in marine limestones. With a geologic record going back to the Precambrian eras more than 500,000,000 years ago, the thallophytes are not only the oldest known plants but also the most persistent and conservative. Examples of widely distributed lime-secreting algae include the rounded, concentrically laminated forms comprising the *Cryptozoon* reefs in the Upper Cambrian limestones near Saratoga Springs, New York. See ALGAE; ALGAE FOSSILS; CHAROPHYTA; CRYPTOZOON

Bryophyta. Among the simplest of the living land plants, the Bryophyta are as soft and perishable as seaweeds and so are equally scarce as fossils. As seen in Fig. 1, the earliest known forms occur in rocks of Pennsylvanian age: *Hepaticites*, a liverwort, from the Coal Measures of England, and *Muscites*, a true moss, from the Carboniferous rocks of France. See BRYOPHYTA.

Psilopsida. This is a little-known group of plants, represented among living plants by the tropical species of *Psilotum*. The Psilopsida are believed to be remotely related to the Psilophytales of Late Silurian and Devonian age. Both living and fossil groups are characterized by their small size, their absence of roots, their leafless stems, and their simple spore-bearing organs borne on short lateral shoots. See PSILOPHYTALES.

Lycopsida. The living forms of this group are the inconspicuous, herbaceous species of *Lycopodium* and *Selaginella*. The Lycopsida include some of the largest, most abundant, and most characteristic trees of the late Paleozoic swamp forests. Beginning in the Early Devonian, the group reached its climax in the Pennsylvanian Period, after which its woody forms declined rapidly and became extinct during the Jurassic Period. Among the best known forms are the tall scale trees (Fig. 2) whose flattened trunks are especially common in the roof shales of coal mines. Their surfaces are marked by closely spaced scalelike leaf cushions and leaf scars, which are diamond-shaped and spirally arranged in the genus *Lepidodendron* and rounded or hexagonal in shape and vertically arranged in *Sigillaria*. See LEPIDODENDRALES, LYCOPODIALES.

Sphenopsida. Plants of this group are characterized by jointed, usually ribbed stems and short, linear leaves arranged in whorls. The only living forms are the so-called horsetail rushes (*Equisetum*) which are essentially small, herbaceous replicas of the tall, woody *Calamites* of the Pennsylvanian coal swamps. None of the arborescent forms survived beyond the Jurassic Period. See CALAMITALES; EQUISETALES; SPHENOPHYLLALES

Pteropsida. Filicineae, the true ferns, have undergone remarkably little change during their long geologic history which extends back to the Devonian Period. The modern ferns (Filicales), so common as underbrush in living forests, can trace

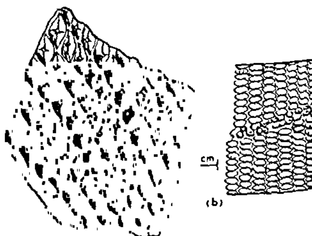


Fig. 2. Scalelike markings on portions of fossilized trunks. (a) *Lepidodendron*. (b) *Sigillaria*.

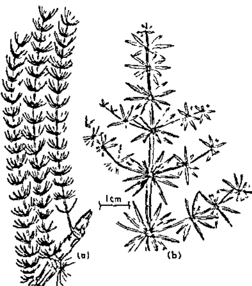


Fig 5 Leafy foliage of the calamites (a) *Asterophyllites* (b) *Annularia*

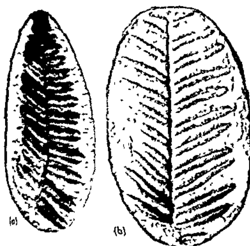


Fig. 6 Pennsylvanian ferns from Illinois showing typical features of leaflets and fronds (a) *Neuropteris* (b) *Alethopteris* (Chicago Natural History Museum)

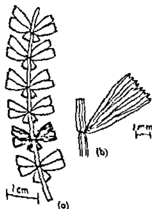


Fig 7. *Sphenophyllum* (a) Branch and leaves (b) Details of leaf venation

zoic (Fig. 1). Greatly reduced in the Permian, the tall, woody lepidodendrons and sigillarias as well as the stately cordaits did not survive beyond this period Also disappearing before the end of the Permian were the ancient ferns (Coenopteridales), the delicate sphenophylls, and the jointed, woody calamites. The seed ferns (Pteridospermae) were severely affected but managed to continue as a minor group into the Mesozoic Era. The conifers, cycads, cycadeoids, and ginkgos, which had barely begun their development at the beginning of the Permian Period, increased in importance during this period of decline in other groups. It is generally believed that the great changes which occurred at this time in both the plant and animal kingdoms were the result of world-wide physical and environmental changes, which also produced a long episode of continental glaciation, mainly confined to the Southern Hemisphere.

Mesozoic floras. Changes in floras are described for the three major subdivisions of Mesozoic time.

Triassic. During the Triassic Period the great transformation in the earth's vegetation, which had begun in the Permian, resulted in the further restriction of the ancient Paleozoic groups of plants and the introduction of many new groups. Ferns continued to be dominant, though most of them belonged to modern families The larger forest trees were mainly conifers cycads, and cycadeoids along with lesser numbers of ginkgos Most of the conifers belonged to extinct genera; a few, including *Araucarioxylon* from the Petrified Forest region of Arizona, were related to groups still living The cycadeoids represented by such well known forms as *Wielandiella* and *Williamsonia*, are completely extinct at the present time The Triassic ginkgos were characterized especially by the deeply dissected leaves of *Sphenobaiera* and *Baiera* (Fig. 8) The few remaining seed ferns were mainly species of *Glossopteris*, *Thinnfeldia*, and *Neuropteridium*.

Jurassic. The vegetation of the Jurassic Period was quite similar to that of the Triassic. Most characteristic are the cycads and cycadeoids, whose broad fronds closely resembled those of the living cycads In the cycadeoids the numerous petrified stems differ from the stems of the cycads chiefly in their possession of "flowers" borne on short lateral branches and largely enclosed by persistent leaf bases and scales. Both the conifers and the ginkgos continued to increase in abundance. Ancestors of both the living pines and redwoods are believed to occur as far back as the Late Jurassic. Mild climates are indicated by the occurrence of typical Jurassic floras as far north as northern Alaska and Siberia and as far south as Graham Land in Antarctica.

Cretaceous In general aspect the floras of the Early Cretaceous were similar to those of the Jurassic Conifers, ferns, and ginkgos continued to flourish, cycads and cycadeoids declined, the latter group becoming extinct by the beginning of Late Cretaceous time Of about 35 genera of conifers known from Lower Cretaceous rocks, 15 are believed to belong to modern groups including the

coal beds which were originally layers of algal debris.

Paleozoic floras. The floras of this era are discussed in chronological sequence from early to late Paleozoic time.

Cambrian and Ordovician. Plant remains continue to be rare in rocks belonging to the Cambrian and Ordovician Periods. Most common are the marine calcareous algae such as *Cryptozoon* from the Cambrian of New York, *Solenopora* from the Ordovician at numerous localities, *Primicorallina* from the Ordovician and Silurian, and the widespread *Girvanella*, which ranged from the Cambrian to the Permian

Silurian and Devonian. Rocks of Late Silurian age in Australia have yielded the oldest reliable record of land plants. These consisted mainly of small, leafless stems (telomes) usually branching dichotomously, and terminated by simple spore-bearing organs, either borne singly (*Sporogonites*) or in clusters (*Hedeia, Yarravia*, and *Zosterophyllum*). Other forms, somewhat resembling modern club mosses, had forked stems covered with short scalelike leaves occasionally bearing small, rounded spore cases (*Baragwanathia*).

In the Early Devonian small primitive land plants belonging mainly to the extinct psilophytes increased in abundance and in geographic distribution. Most characteristic was the genus *Psilophyton* whose small spiny stems (Fig. 4) with terminal paired spore cases are known from localities in eastern Canada, Maine, Wyoming, China, Spitsbergen, and numerous localities in western Europe.

By Middle Devonian time more advanced types of plants made their appearance The lycopods were represented by *Asteroxylon*, whose stems were clothed in short, leaflike scales. The Sphenopsids are known from specimens of thin, jointed stems bearing whorled leaves These belong to the genera *Hyenia* and *Calamophyton* of the extinct order Hyeniales. Fernlike plants, *Protopteridium* and *Aneurophyton*, were simple branched axes with rudimentary frondlike leaves.



Fig 4. Spiny stems of *Psilophyton* from the Lower Devonian of Gaspé.

By the beginning of Late Devonian time one of the major changes in the history of plant life had taken place. The earlier shrubby vegetation, dominated by the primitive psilophytes, had evolved into low forests dominated by woody lycopods and equisetes with underbrush of ancient ferns and sphenophylls. Most abundant in Upper Devonian rocks are the ferns, of which the most widespread and characteristic are species of *Archeopteris, Aneurophyton*, and *Sphenopteris*. In none of these were the fronds and pinnules as well developed as they were in late Paleozoic forms. The lycopods were represented by the scale trees, *Archeosigillaria* and *Cyclostigma*, characterized by small scalelike leaves on dichotomously branched axes. Large, jointed woody equisetes included mainly species of *Archeocalamites* and *Pseudobornia*, the latter bearing whorls of finely divided, featherlike leaves Sphenophylls are of rare occurrence; more common are the petrified stems of the most primitive gymnosperm, *Callixylon*. *See* ARCHAEOPTERIS; HYENIALES; PSEUDOBORNIALES

Mississippian and Pennsylvanian. The development of widespread, dense, lowland forests was attained by gradual floristic evolution in the late Paleozoic Coal Age, referred to in Europe as the Carboniferous Period Prominent in the forests were the stately lycopods, the small-leaved scale trees, *Lepidodendron* and *Sigillaria*, with their relatively unbranched trunks reaching heights of over 100 ft Beside them towered species of *Cordaites* with their long, straplike leaves and lax cones. Somewhat lower in stature were many types of *Calamites* with thick, jointed, ribbed trunks and stems and whorls of linear, pointed leaves (*Annularia* and *Asterophyllites*) (Fig 5). The first true conifers developed by Late Pennsylvanian time; *Walchia* and *Lebachia* are known from numerous impressions of shoots bearing short, closely spaced needles Throughout the Carboniferous the shrubby undergrowth was apparently made up chiefly of ferns and seed ferns (Fig. 6), whose spreading fronds of delicately sculptured pinnules are among the most common fossils found in the roof shales of many of the numerous coal mines developed in the extensive coal beds of this age. Among the most widespread are species of *Neuropteris, Alethopteris, Sphenopteris*, and *Pecopteris*, which are characteristic of the Pennsylvanian of North America and the Upper Carboniferous of Europe Other shrubby plants were the climbers and trailing species of *Sphenophyllum* with thin, jointed stems and whorls of six to nine small, wedge-shaped leaves (Fig 7).

Permian. During the Permian Period many of the great Carboniferous groups of plants began to decline, on their way to ultimate extinction before the end of the Paleozoic E r early in the Meso-

Fig. 10 Angiosperm leaves and fruits Sycamore leaf (a) and fruits (b). (c) Katsura leaf.

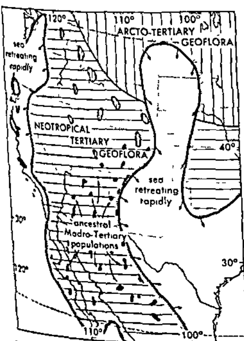


Fig. 11 Inferred distribution of ancestral Madro-Tertiary plants in pre-Eocene time (From D I Axelrod Evolution of the Madro-Tertiary geoflora, Botan. Rev., 24(7) 433–509, 1958)

In summary, migrations of Cenozoic forests toward the Equator have established a trend from trees with large, evergreen leaves to those with small, deciduous leaves This sequence provides a basis for dating plant-bearing rocks with the same degree of accuracy as recognition of first appearances and extinctions in older floras, or as evolutionary development in land mammals. Reconstruction of climate and topography is made possible by comparisons between living trees and their immediate ancestors in Cenozoic rocks [R W C]

Bibliography H. N Andrews, Jr., *Ancient Plants and the World They Lived In*, 1947; C. A. Arnold, *Introduction to Paleobotany*, 1947; D. I Axelrod, Evolution of the Madro-Tertiary geoflora, *Botan. Rev*, 24(7) 433–509, 1958, W C. Darrah, *Principles of Paleobotany*, 1960, A Poldervaart, (ed.), *Crust of the Earth*, Geol Soc Am Spec. Paper 62: 575–592. 1955, A. C. Seward, *Plant Life Through the Ages*, 2d ed, 1933.

Paleocene

The oldest of the five major world-wide divisions (epochs) of the Tertiary Period (Cenozoic Era); the epoch of geologic time extending from the end of the Cretaceous to the beginning of the Eocene, the oldest epoch of the older Tertiary (Paleogene or Nummulitic) *See* CENOZOIC, TERTIARY.



Fig. 8. Leaf of *Baiera*, characteristic of the Triassic and Jurassic Periods.

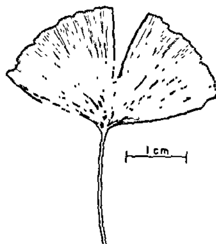


Fig. 9. Leaf of *Ginkgo*, characteristic of the Cretaceous and the Cenozoic, and the only modern survivor of the Ginkgoales.

pines, firs, redwoods, and cedars. The numerous ferns were almost exclusively modern in aspect The ginkgos reached their climax with species belonging to 11 genera, including the sole living survivor, *Ginkgo* (Fig. 9). The cycadeoids are represented by both the typical cycadean foliage and by numerous petrified trunks, especially from the Lower Cretaceous of the Black Hills, South Dakota, and the Chesapeake Bay region.

Of greatest significance during the Cretaceous was the rise to dominance of the flowering plants (angiosperms). The earliest records of the angiosperms actually go back to the Jurassic: pollen showing affinities with both the water-lily and the magnolia families is known from the Middle Jurassic of Scotland; leaves resembling those of the living *Cercidiphyllum*, wood similar to the family Trochodendraceae, and palmlike leaves occur at various Upper Jurassic localities. It is not until the Lower Cretaceous, however, that broad, net-veined leaves, which are believed to be primitive angiosperms, begin to become common. In the Chesapeake region, for example, the oldest Cretaceous beds (Patuxent formation) contain about 6% angiosperms, whereas the overlying units (Arundel and Patapsco formations) have yielded 8 and 40%, respectively, of the leaves of angiosperms (Fig. 10). By Late Cretaceous time angiosperms, including both the Monocotyledoneae and Dicotyledoneae, normally made up 70-90% of the earth's vegetation [E.D.]

Cenozoic floras. The dominant plants during the past 60,000,000 years have been angiosperms and conifers, with ferns, cycads and ginkgos reduced in number. No major groups have appeared or become extinct during this time. However there have been noteworthy changes in distribution, with an equatorward shifting of forest units (geofloras). Cenozoic plants more often occur as impressions than as petrifactions. As a result, attention has been directed toward the distribution of vegetation in time and space, rather than to a study of plant anatomy and phylogeny. The discussion will center in North America, where north-south mountain ranges, in contrast with the predominantly east-west ranges of Eurasia, have provided a terrain favorable for migration with survival.

Most Cenozoic plants are assignable to existing families and genera on the basis of their leaves, with confirming evidence of fruit, stems and pollen. Two floristic units are designated: the Arcto-Tertiary Geoflora dominated by deciduous trees whose nearest living relatives are temperate in occurrence; and the Neotropical-Tertiary Geoflora whose principal members are broad-leafed evergreens like those now living at low latitudes. The Madro-Tertiary Geoflora, characterized by small-leafed shrubs and trees, may have been derived from these in response to progressive aridity in western North America (Fig. 11).

Evidence of wide intercontinental connections is afforded by the uniform composition of the Arcto-Tertiary Geoflora at high northern latitudes during the Eocene. Alder, chestnut, elm, maple, and katsura were forest dominants, with deciduous conifers such as Chinese redwood (*Metasequoia*) more numerous than firs and pines. By Miocene time when this geoflora occupied middle latitudes, new trees were appearing, notably black oaks; several genera which have survived in Asia were becoming rare or extinct in North America. Later Cenozoic forests show further elimination of summer-wet trees from western America, as emergence and orogeny brought continental climate. During the Pleistocene, conifers whose present southern limits are in the Great Lakes region occupied the Gulf states, and the coast redwood ranged into southern California. This represents the climax of the Cenozoic trend toward colder and dryer climate.

The Neotropical-Tertiary Geoflora, best known in the lower Mississippi Basin and from California to Washington during the Eocene, shows a corresponding southward migration At the middle of the era, its palms, laurels, and legumes were still surviving in coastal California. Reduced temperature and rainfall are probable causes for the subsequent restriction of this subtropical forest to its present low latitude occurrence

The boundary between the principal geofloras during the Cenozoic bends southward across the northern continents from west to east and turns northward over the oceans, as do existing isotherms. Throughout the era these primary relief features appear to have controlled the pattern of temperature and vegetation as they do today, a relationship which offers substantial evidence against the theory of continental drift during later geologic time.

ctions, as well as the annual variations among ese As the information and interest in present imate are concerned principally with the lands, nalogous data for the past should be gained by bserving land laid sedimentary rocks and the soils ormed on ancient land surfaces. The principal or-anic remains in terrestrial sediments are the lants though vertebrate animals and a few classes f invertebrates are locally useful. Though marine climates are of less concern in the present, there is abundant information on past marine conditions, for it is in the seas that the sediments are largely preserved Marine rocks give some indications of climates on source lands, because the rocks there were affected by weather conditions.

Terrestrial sediments. Terrestrial sediments are deposited by streams (fluviatile), and in lakes (lacustrine), or by the wind (aeolian) The directions

...

es of mountain ranges. Winds in areas of low ecipitation may cause sand dunes to drift; the rection of drift is determined from the cross ratification of beds deposited on the lee slopes of ich dunes *Orientation* studies have been made in any regions but have given no more than provin-al knowledge of winds

The petrology of *terrestrial* sediments indicates hether their components were assembled under onditions of *temperature* and moisture that altered ss stable minerals and converted them to more able types. Sediments rarely preserve soil zones hat can be compared to those on present surfaces, uch information has been applied particularly to Pleistocene glacial deposits Temperature and pre-cipitation characteristics at the site of deposition are reflected by the *cementation* of the sediment and the presence of such substances as alkali salts that indicate relatively ... of oxidati... environme... is interre... to the sed*imentary* petrology, for oxidation or reduction depends *in* many instances on the abundance of plants, which in turn reflects the temperatures and precipitation *See* SEDIMENTATION (GEOLOGY)

Seasonal variations in temperature are revealed in varves, repetitions of winter and summer layers, such as those in lake deposits that show the seasonal abundance of *organic* matter and those in glacial lake clays that show the influences *of summer* melting and winter freezing of glacial *ice on* the transporting power of streams that feed the lakes. See VARVE

Marine sediments. Marine environments are more constant *in* annual temperature range However, some marine sediments *were* laid in lagoons or seas having restricted water inflow and high evaporation, which led to the *deposition* of salts Common salt, or sodium chloride (halite), precipitates from normal sea water when it is concentrated to less than one-tenth of its *original* volume, and other salts precipitate at lower and higher concentrations. The nature of the precipitated substances can depend on the temperature, hydrogen ion concentration (acidity-alkalinity), redox potential (oxidation-reduction), and ion composition of a marine water The form and clarity of halite crystals are related to water temperatures; hence salt deposits may show seasonal banding. Some substances are of different compositions at different temperatures; thus calcium sulfate is in anhydrous form (anhydrite) at higher temperatures than the hydrous mineral (gypsum). Other substances, such as calcium carbonate (calcite and aragonite), have different crystal forms under different temperatures and other factors A geologic thermometer of promise is found in the ratios of oxygen isotope 18 to isotope 16, which vary with temperature of crystallization in calcium carbonate The average temperature of the sea can be determined from study of fossils, and analysis of carefully separated annual layers can yield temperatures of summer and winter seasons *See* GEOLOGIC THERMOMETRY; MARINE SEDIMENTS, PALEOECOLOGY (GEOCHEMICAL ASPECTS).

Invertebrate fossils provide the best source of information about marine environments. Reef corals indicate low latitudes and warm temperatures by analogy to those in present coral reefs Although many organisms have limited ranges of temperature and salinity tolerance, marine conditions are so universally temperate in some geologic periods that climatic belts have not been recognized.

Climatic zones. Few world climatic maps have been prepared from the many possible sources of information There is disagreement with regard to the stability of climatic zones relative to the present geography. Some believe that the climatic belts of the Tertiary, of the Jurassic, and of the Devonian, Carboniferous, and Permian were generally accordant with those of today, paralleling *the* Equator, with poles near their present positions. Others have maintained from climatic evidence alone that late Paleozoic and Jurassic poles were quite differently placed than at the present The distribution of climatic zones in the past and the studies of paleomagnetism (the determination of magnetic poles from remanent magnetism in ancient rocks) are the two means by which the positions of the poles of earth rotation can be determined *See* ROCK MAGNETISM.

Paleomagnetism gives measurements of some precision and gives longitudes as well as latitudes The magnetic poles and poles of rotation seem to coincide on the basis of statistical averages *over a* span of time, even though the present magnetic poles are rather far from the North and South Poles If the poles have moved in the past, *climatic* zones must have deviated from parallelism with present latitudes The two sorts of evi*dence bear* on continental drift. for the rock magnetism as well as the evidence of climatic latitude*s in the* rocks must fit an integrated map If latitudes and *longitudes* changed significantly for different continents in the

Nomenclature. Sir Charles Lyell, the British geologist, in 1833 subdivided the Tertiary into Pliocene (youngest), Miocene, and Eocene. Subsequently, as a result of studies of fossil plants, W. Schimper in 1874 introduced the term Paleocene because he believed the plants in the basal part of the Eocene were sufficiently different from those higher in the Eocene to justify their separation. The term did not achieve much popularity as a division of the Tertiary until about 1920, but it has now come to be rather widely used in many regions of the world. Since approximately 1950, certain beds which had previously been classed as latest Cretaceous (Danian of western Europe and elsewhere) have been generally recognized as Paleocene.

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

Tertiary					Quaternary	
Paleocene	Eocene	Oligocene	Miocene	Pliocene	Pleistocene	Recent

The Paleocene Series includes all rocks formed during the Paleocene Epoch, but the term is used most specifically for the sedimentary rocks formed during this portion of the Tertiary Period These strata contain the plant and animal remains which are the primary bases for the identification of Paleocene age.

Although the limits of the Paleocene are marked in places by physical breaks in the rock record, such physical changes are not present everywhere, and accordingly, differentiation of Paleocene from Cretaceous beds below and Eocene beds above is often difficult. In these instances identification of Paleocene, Cretaceous, and Eocene is primarily a comparative paleontologic problem

Strata. The Paleocene strata tend to be relatively similar to the underlying Cretaceous, but they also consist of all the common sedimentary types, varying from marine through marginal-marine or intermediate to terrestrial. They are typically unconsolidated and are widely dispersed throughout the world. The terrestrial beds are best known in the continental interiors, while the marginal and marine beds are most common near the continental borders in the areas of the coastal plains and continental shelves Especially noteworthy examples of these various strata are known in (1) the Gulf Coastal province of the United States and Mexico, (2) the intermontane basins of the North American Cordillera; (3) northwestern Europe; (4) the Mediterranean Sea region of southern Europe and northern Africa; and (5) Asia Minor and parts of India.

Most of the known marine and marginal strata are relatively undisturbed and flat-lying and occur [illegible]

sea level. Still other Paleocene strata have been depressed or have subsided below sea level. The terrestrial strata were deposited above sea level in the continental areas, and, for the most part, have remained in this position. In general it may be said that the sedimentary rocks of this age are more deformed than younger strata but are less modified than older rocks. Known thicknesses of Paleocene deposits vary from a few feet to thousands of feet. Generally speaking, the Paleocene rocks are not as important economically as other Tertiary rocks, but in various places they have yielded important quantities of oil and gas, fresh water, clay, sand, marl, limestone, and other products.

Igneous rocks of Paleocene age include both intrusive and extrusive varieties, the latter being best known because they commonly occur as layered volcanic materials at the earth's surface and are therefore more readily available for observation Igneous activity, generally, was not as widespread in the earlier Tertiary (Paleogene or Nummulitic) as in the later Tertiary (Neogene); consequently Paleocene igneous rocks are not ordinarily as abundant as those of Miocene and Pliocene age

Fauna. The Paleocene terrestrial fauna is dominated by archaic mammalian types (marsupials, insectivores, creodonts, and condylarths) from which the more advanced carnivorous (cats, dogs, bears) and herbivorous (horses, cattle) mammals evolved in Eocene time and later A specialized tropical marine fauna, including characteristic pelecypods, gastropods, echinoids, and foraminifers, appeared in the Indo-Pacific region in Paleocene time and started to spread westward toward the Mediterranean See PALEOBOTANY; PALEONTOLOGY. [A.H.CH.; C.E.M.]

Paleoclimatology

In geology, that part of paleogeography that concerns temperatures, precipitation, winds, and other weather conditions and their zones of distribution The interpretation of past climates is based on the compositions and structures of sedimentary rocks and on their organic remains. In addition to the intrinsic value of information concerning the past history of the earth's surface, knowledge of climatic zones through time has significance in connection with questions pertaining to the stability of the crust with reference to the axis of rotation, and the permanence in relative position of continental and oceanic areas. See PALEOGEOGRAPHY.

Present climatic belts are distinguished by their temperatures, rainfall an[illegible]l, and wind di-

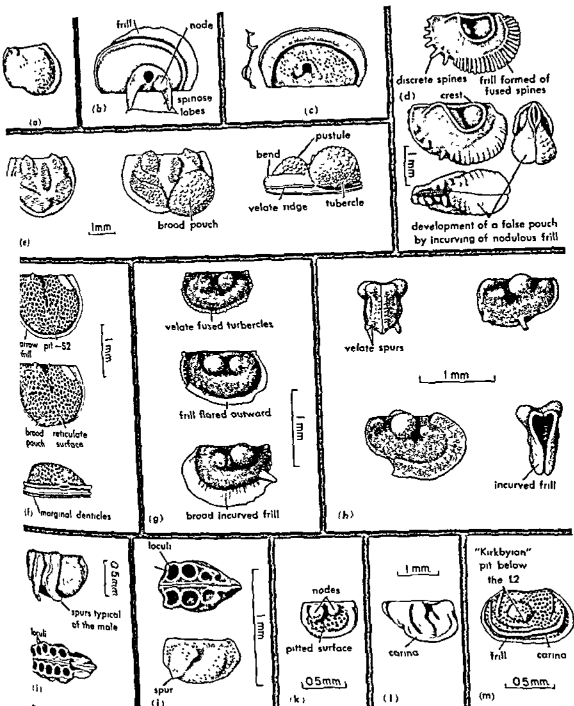


Fig. 2. (a) *Dihogmochilina latimarginata* (Jones), left valve of a Silurian leperditacean ostracod. (b) *Eurychilina subradiata* Ulrich, right valve and cross section (c) *Dicranella bicornis* Ulrich, left valve (d) *Piretella acmaea* Öpik, male right valve and female carapace (e) *Beyrichia tuberculata* (Klöden), male and female right valves. (f) *Hibbardia lacrimosa* (Swartz and Oriel), male and female right valves (g) *Hollinella dentata* Coryell, immature left, male right, and female left valves (h) *Falsipollex laxivelatus* Kesling, male and female carapaces. (i) *Ctenoloculina cicatricosa* Warthin, male and female carapaces (j) *Tetrasacculus mirabilis* (Croneis and Gale), female and male carapaces (k) *Ulrichia affinis* Swartz, left valve. (l) *Sigmoopsis platyceras* (Öpik), male left and female right valves (m) *Amphissites marginiferus* Roth, left valve.

tion, Middle Ordovician ostracods of central and southern Sweden, *Bull Geol Inst Univ Upsala*, 37:173–442, 1957; R V. Kesling, Terminology of ostracod carapaces, *Contrib Museum Paleontology Univ. Mich.*, 9(4):93–171, 1951, P. C. Sylvester-Bradley, The structure, evolution and nomenclature of the ostracod hinge, *Bull Brit. Museum Geol.*, 3(1), 1956.

past, paleoclimatologic maps should be made on base maps unlike those of the present geography.

[M.K.]

Paleocopa

A suborder of extinct ostracods with straight hinges of some length (in most genera). If the Ostracoda are considered to be a subclass, then the Paleocopa could be considered an order. No frontal opening is present and little or no duplicature, or calcareous inner lamella, occurs. Lobes and sulci are conspicuous in many families, and are conventionally designated from the anterior end of each valve, as L1, S1, L2, S2, L3 (Fig. 1). By homology with quadrilobate ostracods, S2 is the most conspicuous sulcus and is present in all lobate forms; it marks the position of the internal adductor muscle scar. The soft parts are unknown The suborder is divided into two superfamilies, the Leperditacea and the Beyrichiacea. *See* OSTRACODA.

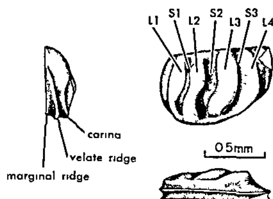


Fig 1. *Ogmoopsis nodulifera* Hessland, left valve with lobes and sulci labeled

Leperditacea. This Ordovician-Devonian superfamily is characterized by a long hinge, greatest height and greatest width posterior, sulci weakly developed or absent, and by smooth, unornamented valves Many species have an eye tubercle. Leperditacea contains the largest known ostracods *Dihogmochilina gigantea* (Roemer) is reportedly 43 mm long, and *D. latimarginata* (Jones) exceeds 30 mm.

Beyrichiacea. This Ordovician-Permian superfamily includes ostracods of many kinds Several families exhibit dimorphism in the adult carapaces, such as Beyrichiidae, Hollinidae, Sigmoopsidae, and Kloedenellidae. In contrast, many have only one adult form, such as Eurychilinidae, Drepanellidae, Kirkbyidae, and others.

Monomorphic families. Eurychilinidae (Ordovician-Silurian) had a frill from corner to corner and a pit, or short sulcus, for S2. Some species, like *Eurychilina subradiata*, developed a broad incurved frill. Through genera like *Dicranella*, they may be ancestral to the Piretellidae (Ordovician), in which the male has a broad frill and the female a false pouch (*Piretella*) formed when the two frills incurve to meet along their outer edges.

Drepanellidae (Ordovician-Carboniferous) is a nondimorphic family having a basic lobation of two concentric U-shaped ridges In some genera, like *Ulrichia*, however, the middle ridge is reduced two nodes.

Kirkbyidae (Carboniferous-Permian) has non morphic species with frills, reticulate surfac and a subcentral pit known as the Kirkbyian p *Amphissites* is a representative genus with a we developed carina. The closely related and probab ancestral family Arcyzonidae (Devonian) has large central pit and less conspicuous carina.

Aechminidae (Ordovician-Carboniferous) co tains ostracods with weak lobation and a promine hollow centrodorsal spine. In some species *Aechmina*, the spine is much larger than the re of the valve Its use is unknown

Dimorphic families. Beyrichiidae (Siluria Devonian) probably evolved from the Piretellid when the female frill incurved to join the ventr edge of the valve; the body wall, so enclosed, di *appeared; and a pouch opening into the interi* developed. Thin sections reveal very young instar or developmental stages, of the species within the pouches, and confirm that they were brood space In Silurian beyrichiids, like *Beyrichia*, pouch were bulbous and anteroventral, but in the la Devonian survivors, like *Hibbardia*, they we smaller and posteroventral.

Hollinidae (Ordovician-Permian) has dimo phism in the velate structure. Most genera have *knoblike L3 In Hollinella, the young have a ro* of tubercles or a narrow ridge, the male has a na row frill flared outward, and the female a wide in curved frill. In *Falsipollex*, the male has velat spurs and the female an incurved frill *Ctenolocu lina* is quadrilobate, with spurs at the ends of L1 L2, and L3 in the male and a scalloped frill an pocketlike loculi, curious unexplained structures in the female *Tetrasacculus* is bilobate, with a anteroventral spur in the male and large loculi i *the female Other hollinid genera* have differen combinations of lobation and velar dimorphism.

Sigmoopsidae (Ordovician) is characterized b carinal dimorphism, as shown in *Sigmoopsis*. Basically, valves are quadrilobate, but in some genera S1 and S3 are weak or absent.

Kloedenellidae (Devonian-Permian) shows dimorphism in the posterior region, which is rather narrow in the male and large and inflated in the female Most genera exhibit conspicuous overlap, and many have terminal notches on the hinge

Kloedeniidae (Silurian-Devonian), closely related to Beyrichiidae, shows an anteroventral swelling in the female instead of a well-defined brood pouch *See* SEXUAL DIMORPHISM

Not all paleocopan families are discussed here. There are differences of opinion concerning the taxonomic position of members of the group which arise from considerations of relative importance of dimorphism versus lobation, and from interpretation of structures. [R V.K.]

Bibliography: R S Bassler and B. Kellett, *Bibliographic Index of Paleozoic Ostracoda*, Geol. Soc. Am. Spec. Paper 1, 1934; H. V. W. Howe, *Handbook of Ostracod Ta*[illegible], Louisiana State Univ. Studies, Phys. Sc[illegible]; V. Jaanus-

known as inner neritic facies Paleoecolo- empt to recognize comparable facies in the :ous rocks that make up a large part of the :rust (Fig 1). *See* DEEP-SEA FAUNA, ZOO- HY

)up of organisms that live together as a ;ity is termed a biocenose by ecologists, eful term cannot be carried over directly eoecology for two reasons: (1) many of the ms that made up an important part of a)mmunity do not lend themselves to fossili- and are rarely preserved; and (2) before mblage is buried and thus fossilized the composition may be radically altered by lition of shells or other skeletal parts of and plants that lived elsewhere under dif- conditions A fossil assemblage

t sediments because the extraneous elements ined must be recognized and eliminated

ure of evidence. Some paleoecological inter- m the rocks from

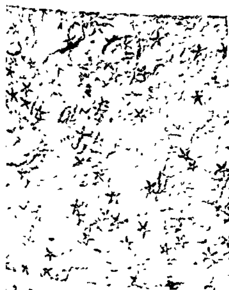
...ces are used

a from fossils The shape of a cephalopod or other swimming or crawling form may sug- t mode of locomotion, the thin and fragile of certain foraminifers or gastropods may clearly to a pelagic existence Corals and tary or burrowing pelecypods and brachio- may be found in position of growth, indicating they probably were buried in the place where they lived When the members of a fossil assemblage are compared with living relatives whose life habits are known (Fig. 2), it may become apparent that the fossil assemblage is a mixed one with representatives from more than one environment. Such an interpretation may be confirmed if some members of the group exhibit signs of breakage or wear or if only the lighter unattached valves of sedentary forms are present Studies of the isotopic composition of the shells may indicate the temperature or some other aspect of the past environment. *See* PALEOECOLOGY (GEOCHEMICAL ASPECTS)

Many of the older rocks contain fossils that have no close living relatives Among such extinct groups are the conodonts, the archaeocyathid "corals," and the graptolites. Speculation about the conditions under which such organisms lived may, of necessity, be based largely on the supposed living habits of associated organisms and partly on the lithology and structures of the sediments containing the fossils *See* ARCHAEOCYATHA; CONODONT, GRAPTOLITHINA

Knowledge of the living habits of existing organisms is not always a sure clue to the habits of ancestral types. Today, for example, the stalked crinoids are solitary forms found at great depths but their numerous ancestors in Paleozoic times were gregarious and inhabited shallow waters (Fig 3)

Data from rocks The rock containing the fossils may indeed be the major source of interpretative data The texture of the enclosing sediments may reveal much about the site of deposition In water-laid sediments the occurrence of graded beds, that is, beds in which the texture grades upward from coarse to fine, may point to landslides or some form of turbidity current that threw the sediment into



. 2 Devonian sandstone showing starfishes and ... It has been suggested that the starfishes were ...ing on the clams when buried Width of area shown ...ut 18 in. (Photograph from D W Fisher, New York ...te Museum)



Fig 3 Gregarious, free-swimming crinoids (*Uintacrinus socialis*) from the Cretaceous rocks of Kansas Rounded cups are 1½–2½ in in diameter (Photograph from H S Ladd, ed, Geol Soc Amer Mem. 67, vol 2, 1957, courtesy of Smithsonian ...)

Paleoecology

The ecology of the geologic past, a study of the relations of fossil organisms to each other and to the environments in which they lived. Paleoecology is a study based on inferences and interpretations and on the basic assumption that the animals and plants of the past lived under essentially the same environmental conditions as do living relatives

Scope and aims. As in studies dealing with living organisms, all types of environments are considered in paleoecology: marine, brackish, and fresh waters and land areas. Marine environments have received a major share of attention because most of the fossils that make up the paleontological record are contained in sediments that were deposited in the sea. *See* ECOLOGY

The aim of paleoecology is to infer, in terms of present-day conditions, the physical environments in which fossil organisms lived and their relations to each other in the changing environments of the past. In studying the record as preserved in the rocks, the paleoecologist attempts, as does the ecologist, to make a complete list of the animals and plants that inhabited a particular area and to obtain an estimate of their relative abundance. This is a difficult task even for the ecologist For example, to compile such a census along a given stretch of coast today is not easy, because the most intensely studied shores yield new occurrences as investigations are continued The chief cause for this is the rarity of some species, but there are other difficulties: some species burrow far below the surface and others appear only at night or at certain seasons of the year. The comparable task of a paleoecologist is still more difficult He has access only to a limited amount of sea bottom and in the beds that were laid down at a particular time the remains of only a small fraction of the population then living were preserved Furthermore, this meager record may contain a mixture of forms, including some that lived at the spot of burial and some that were brought in from another and perhaps quite different environment.

The first written observations on the subject now recognized as paleoecology were made by the Greeks about 500 B.C., but these were few and widely scattered. What might be called modern ecology did not begin to take form until the early part of the eighteenth century and paleoecology began to develop at about the same time

Terms. Studies of the distribution of living plant and animal life, both on the land and in the sea, have led to the recognition of communities or assemblages, each of which is adjusted to the physical and biological conditions that make up its environment. In the sea, for example, the assemblages of animals and plants that live at intertidal levels differ from those found at greater depths These fairly obvious distinctions were recognized before 1860 and names were given to major bathymetric and biogeographic zones The major subdivisions of the sea are divided into smaller units All such units are referred to as facies Thus a deep-water assemblage is referred to as a bathyal facies, and an assemblage from the inner shelf

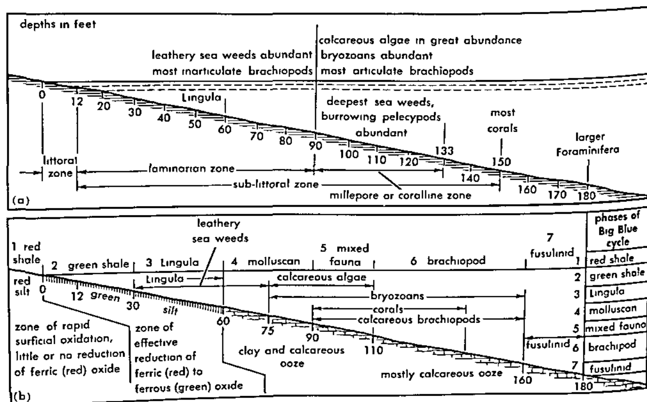


Fig 1 Sea-bottom zones in existing and ancient seas. (a) Ideal distribution of benthonic organisms in shallow waters of existing seas (b) Restoration of the sea-bottom zones in Big Blue (Permian) time in Kansas. (*After M K. Elias, Depth of deposition of the Big Blue, Late Paleozoic, sediments in Kansas, Bull. Geol Soc Amer, 48(3).403–432, 1937*)

the coast of California With the aid of subsidence, many thousands of feet of beds were deposited Large parts of the section are fine-grained shales and siltstones whose contained Foraminifera (when compared with living forms collected from known depth zones) indicate depths at which deposition occurred as great as 4000 ft. Sandstones alternate with the fine grained sediments and many of these are graded, indicating that they probably were transported into the basin by turbidity currents. Beds of gravel and massive conglomerate also occur and these are believed to have been emplaced by landslides. During parts of the Pleistocene the basins were filled with marine sediments, and terrestrial sediments containing the remains of land vertebrates were laid down *See* SEDIMENTATION (GEOLOGY).

Other examples involve specialized types of sediments such as black shales and limestones. Ecologists have described many areas in which black muds are being deposited today, in which there is little or no circulation, in which the supply of oxygen is low, and consequently, in which there is practically no benthonic (bottom-dwelling) life Many of the black shales of the geologic column record deposition under such conditions in past times. R. Ruedemann has cited specifically the dark shales that are widely developed in eastern New York and concluded that planktonic (floating and weakly swimming) organisms were brought in freely by currents, but that at deeper levels circulation was poor and toxic conditions impoverished or prevented the existence of benthonic life. *See* BLACK SHALE.

An environment comparable to that of the black shales described by Ruedemann is recorded in the La Luna limestone and its equivalents, which constitute a unique lithologic unit 500–3000 ft thick widely developed over much of northern South America and some nearby regions in Cretaceous time. H. Hedberg has showed that the La Luna, a dark carbonaceous limestone, is composed almost entirely of the tests of pelagic Foraminifera and that large fossils are rare. He expressed the opinion that at the time of deposition the sediment was a *Globigerina* ooze and speculated that life in the La Luna seas was almost exclusively planktonic. Thus, according to this theory, the bottom waters developed a toxicity (because of lack of circulation) that rendered them uninhabitable by marine benthonic animals. Such conditions would permit the accumulation of dead plankton on a sea floor undisturbed by bottom scavengers The toxicity and the lack of oxygen were thought to have prevented rapid bacterial decay and permitted the accumulation of soft organic matter Much of this material is believed to be still present in the form of carbonaceous-bituminous matter with some free petroleum. The rock appears in many ways to be an ideal petroleum source bed. *See* FORAMINIFERA FOSSILS, *see also* LIMESTONE.

Relations to other sciences. The interpretations made by the paleoecologist have direct applications to other fields of earth science, especially paleontology and stratigraphy In paleontology, for example, one of the problems facing the investigator is that of determining the significance of concentrations of fossils on individual bedding planes in many parts of the geologic column. At Lompoc, California, the Miocene shales contain a bed the surface of which is covered with skeletons of a species of herring (Fig 5) This bed once formed



Fig 5 Catastrophic death in the Miocene Skeletons of herring preserved on a bedding plane of diatomaceous earth in the Monterey shale of Lompoc, California Skeletons are 6–8 *in long* (Photograph from A B Cumings, Johns-Manville Co, in H S Ladd, ed, *Geol. Soc. Amer Mem 67, vol 2, 1957*)

suspension before permitting final settlement. The presence of certain types of ripple marks, mud cracks, rain prints, or other sedimentary structures may aid in determining the nature of the original environment. Other interpretations may be based on a high content of organic matter, lime carbonate, or on the presence of oolites, glauconite, or nodules of phosphorite or manganese oxides (Fig 4). The occurrence of such materials is significant because limnologists and oceanographers have studied the areas in which they are being formed today. *See* SEDIMENTARY ROCKS.

Data from fossils and rocks. Combinations of the above-mentioned types of data may give a suggestion or a clear indication, in the case of water-laid sediments, about such features as type of bottom, nearness to land, depth, agitation, and tur bidity. Specific examples to illustrate this type o interpretation are given below.

Common sediments such as mud, silt, sand, an gravel are deposited under a variety of conditions When these deposits are elevated at a later geo logic time as shale, siltstone, sandstone, or con glomerate, the exact type of the original sedimen may mean very little except to indicate that th lake or sea bottom at the time of deposition wa muddy, sandy, or covered with boulders. Howeve when the beds in a given elevated section sho variable texture, they preserve a record of changin times, possibly a record of unusual events. Suc records are preserved in sediments deposited du ing Tertiary and Pleistocene times in basins alon

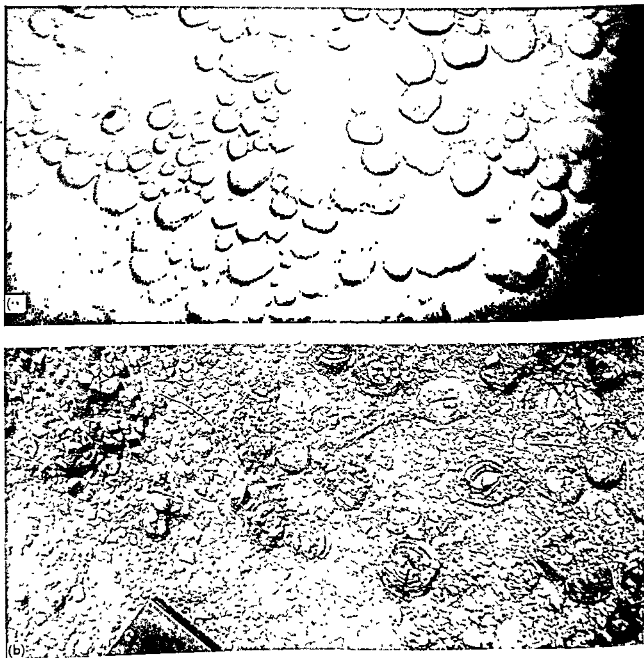
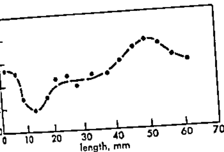


Fig 4. (a) Manganese nodules at a depth of 3000 fathoms in the North Atlantic. Largest nodules are 5 in in diameter (*photograph by D M Owen, Woods Hole Oceanographic Institution, from H S Ladd, ed, Geol. Soc. Amer. Mem. 67, vol. 2, 1957*). (b) Manganese nodules in red deep-sea clay of Cretaceous age on Island of Timor, Indonesia (*photograph by H G Jonker, courtesy of J G. Ubaghs, Mineralogisch Geologisch Museum, Delft, from H. S Ladd, ed, Geol. Soc Amer. Mem. 67, vol. 2,* .



1 Cyclic changes in aragonite-calcite ratios for ...ecutive growth increments of a worm tube from ...muda inshore waters

...h as those of the calcareous tubes of serpulid ...orms. Figure 1 shows the aragonite-calcite ratios ...r consecutive growth increments of a worm tube ...llected from the Bermuda inshore waters where ...e yearly temperature variations are large, rang...ng from 16 to 30C°. The specimen was taken ...live in early summer prior to the seasonal maxi...mum of water temperatures. The aragonite calcite ratios are plotted against the length of successive growth increments, with the last increment shown at the left of the diagram The effect of temperature is shown by the cyclic changes in the aragonite-calcite ratios. Summer growth zones are 90% aragonite, whereas the minimum value for winter growth is 60% aragonite.

The physiology of the organisms also appears to influence the aragonite-calcite ratios Individuals of different species grown in the same temperature may show differences in aragonite-calcite ratios A more limited effect of temperature on the skeletal mineralogy is observed in certain gastropods and pteropods which range from tropical or subtropical to temperate waters The shells of these mollusks consist of 100% aragonite in all but those that live in the coldest waters Traces of calcite, [illegible] ...est tem- [illegible] certain [illegible] ...ate precipitation on temperature These aragonite-precipitating species are essentially limited to the tropical belt bounded by the 16°C isotherms for the coldest month of the year Evidently the aragonite is a passive warm-water precipitate

The aragonite calcite ratios may also depend upon the water chemistry. This is indicated in species which range from mean ocean to brackish waters The few aragonite-calcite ratios determined for individuals from waters of low salinity appear to be noticeably higher than the ratios for individuals in the same temperature regime from normal marine waters

Magnesium content. The Mg concentrations of calcareous skeletons in Recent species depend on several factors: the mineral form, the phylogenetic level of the species, the water temperature, and the water chemistry. The calcitic structure can accommodate a considerably larger amount of Mg in solid solution than can the aragonitic structure Biological aragonite precipitates therefore rarely have over 1 mole % $MgCO_3$, whereas the calcitic ones may contain as much as 19 mole %.

In calcitic skeletons of individuals within the same phylogenetic class the $MgCO_3$ content shows an increase with higher environmental temperatures. The relative concentrations of Mg and the slopes of the magnesium-temperature curves differ for different classes, as shown in Fig 2. The slopes of the magnesium-temperature curves tend to become lower as the phylogenetic rank of the class becomes higher. However, data on an echinoid species indicate that the magnesium-temperature curve for a single species may deviate from the curve for the class as a whole.

The magnesium content of carbonate skeletons is also affected by changes in the chemistry of the sea water. In the same temperature regime, echinoids from hyposaline waters have lower magnesium contents, and articulate brachiopods from hypersaline waters have higher magnesium contents than conspecific forms or related species from mean ocean waters.

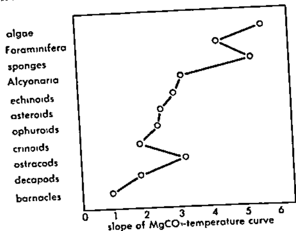


Fig 2 Relation between slope of temperature-magnesium curve and organic complexity

Strontium content. The Sr content of the calcareous skeletons in present-day species is affected by the same factors which influence the Mg content The effect of crystal form, however, differs from that of Mg in that Sr is more readily accommodated in the aragonitic than in the calcitic structure The $SrCO_3$ content of calcitic skeletons rarely exceeds 0.4 mole %, whereas in aragonitic skeletons its content may be as high as 1.3 mole %

An effect of temperature upon the Sr content in calcareous skeletons has been demonstrated so far only in two groups of organisms, the articulate brachiopods and the echinoids The crystal form in both groups is calcite. Figure 3 shows the relationship between Sr content and temperature in articulate brachiopods The temperature values are based on determinations of the O^{18}/O^{16} ratios of total shells, corrected for the O^{18}/O^{16} ratios of the waters from which the specimens were derived. The $SrCO_3$ content is shown to increase with higher environmental temperatures The curve is based on data from samples of species from several super-



Fig 6. Catastrophic death in the sea today. Fishes killed by red tide in the Gulf of Mexico, 5 miles south of Sanibel Island, November, 1953. (Photograph by Kenneth Marvin, U.S Fish and Wildlife Service, fro H S Ladd, ed., Geol. Soc. Amer. Mem 67, vol 1957)

the bottom of a bay comparable to those along the coast today. In the bay, over an area of 4 mi², more than 1,000,000,000 herring died almost simultaneously. Ecological studies show that catastrophic death on a comparable scale occurs in the sea today, brought on, in many instances, by the appearance of red water or the red tide (Fig. 6). The upwelling of cold waters along a coast brings a rich supply of nutrients to the surface and this may lead to the development of noxious blooms of microscopic flagellates and dinoflagellates fatal to fishes. Ecological studies show that such red water is only one of several causes that may bring about mass mortality in the sea today and may have been responsible for the preservation of crowded layers of fossils in the past Other causes include volcanic eruptions, tidal waves, and rapid changes in temperature and salinity

Paleoecology is closely allied to stratigraphy, the branch of geology dealing with the sedimentary rocks and their order of superposition (chronological sequence) The paleoecologist, familiar with the distribution and living habits of life in existing environments, may be able to show, for example, that dissimilar faunas in two separated geologic sections may be contemporaneous and that two assemblages that resemble each other may merely reflect a similar environment, one being appreciably older than the other *See* STRATIGRAPHY.

Paleoecological interpretations assist the geologist in reconstructing ancient landscapes, shorelines, and sedimentary basins For this reason they are of value to those engaged in the search for oil and gas The same is true in investigations of mineral deposits, especially those that have a stratigraphic control, such as phosphate and associated trace elements. *See* PALEOGEOGRAPHY. [H.S.L.]

Bibliography: C. O Dunbar and J. Rodgers, *Principles of Stratigraphy*, 1957; H. D Hedberg, Cretaceous limestone as petroleum source rock in north-western Venezuela, *Bull. Am. Assoc. Petrol. Geologists*, 15(3):229–246, 1931; J. W. Hedgpeth (ed.), *Treatise on Marine Ecology and Paleoecology*, Geol Soc. Am. Mem. 67, vol. 1, 1957; G. E. Hutchinson, *A Treatise on Limnology*, vol. I, 1957; W C Krumbein and L. L. Sloss, *Stratigraphy and Sedimentation*, 1951; H. S. Ladd (ed), *Treatise on Marine Ecology and Paleoecolog* Geol. Soc. Am. Mem. 67, vol. 2, 1957; H Lad Ecology, paleontology, and stratigraphy, *Scienc* 129(3341):69–78, 1959; R. Ruedemann, Ecolog of black mud shales of eastern New York, *J. Paleontology*, 9(1):79–91, 1935

Paleoecology (geochemical aspects)

The study of the chemical and mineral compositic of fossil organisms as these relate to the ecolog cal and geochemical features of ancient enviro ments. Chemical studies of fossils, particular those that are abundant and widespread and th have related modern forms restricted either marine or to fresh-water environments, yield val able evidence concerning the paleoecologic ar evolutionary aspects of skeletal building material For further treatment of the diagnostic use of fo sils, *see* FOSSIL, PALEOECOLOGY, *see also* PALE BIOCHEMISTRY

Chemical studies on the carbonate skeletons Recent marine organisms have shown that the ter perature and chemistry of sea water in which tl organisms grow may influence the contents strontium, Sr, and magnesium, Mg; the ratios oxygen-18 to oxygen-16 (O^{18}/O^{16} ratios); and tl aragonite-calcite ratios of the skeletons. Speci information about the relationships between tl two ecologic factors and the characteristics of tl skeletons is given in examples on studies done present-day organisms and is followed by certa paleoecological applications of the data.

Skeletal mineralogy. In recent species of ce tain bryozoa, serpulid worms, pelecypods, and ga tropods, the skeletons are commonly composed aragonite and calcite The two mineral modific tions always form separate microarchitectural uni of the skeletons. The aragonite-calcite ratios for th skeletons as a whole among individuals of a give species in a given environment vary with the age the individual The variations in the aragonit calcite ratios within a species may be correlate with temperature; the higher the temperatures, th higher is the aragonite-calcite ratio. The depen ence of the aragonite-calcite ratios upon temper ture is most clearly defined in individual skeleton in which skeletal gr [illegible] periphera

tions of lands and seas, their elevations, depths, and forms. A number of analogous matters commonly are treated as a part of paleogeography, though really stratigraphic, involving spans of time during which a thickness of sediments was laid. See STRATIGRAPHY.

Time and correlations. Paleogeography generally involves the correlation of the rock record and events at the time being considered. In a local area or province, the criteria may be lithic, inasmuch as conditions may have been such that a single kind of sediment was synchronously deposited throughout a considerable area of land or sea. The identification of time is substantiated by study of organisms, for fossils are the principal means of carrying time correlations among distant places. Isotope ratios among uranium-lead, thorium-lead, potassium-argon, strontium-rubidium and carbon isotopes are used increasingly in varying degrees of accuracy for differing spans of time. The problems of correlation are in the domain of stratigraphy, but they are basic to paleogeography. *See* ROCK (AGE DETERMINATION).

Map projections. The preparation of paleogeographic maps involves not only the same problems of distortion of shapes, forms, and areas that are encountered in all map projections, but there are additional difficulties. Most available geologic maps fail to present the proper relative positions of the rocks at the time that the paleogeographic map portrays. When rocks are deformed by folding or faulting after their deposition, the positions and directions between points on opposite sides of each fold or fault have been changed. If a continent has moved relative to another, or a great mountain range has risen through compression, significant deviations develop in the subsequent and present geography from that prior to the deformation. Maps that reconstruct the original relative positions of rocks are known as palinspastic maps. Generally the limits of error in the paleogeography are so great that the present geographic base maps are reasonably satisfactory.

Paleogeographic maps. The simplest forms of paleogeographic maps show the distribution of lands and seas. The study of sedimentary petrology permits determination of the nature of the source lands of the time. Orientations of depositional structures such as ripple marks, cross stratification, flow casts, and elongations of particles and organisms reveal directions of streams, currents, and winds, aspects of paleoclimatology. The elevations of lands can be shown approximately by hypsometric contours of elevation, with colors or shades as used for present geography; more often lands are shown by standard geomorphic landform symbols that give better impressions of the character of lands and less emphasis on elevations. Paleogeologic maps, showing the pattern of rocks at the surface at a past time, aid in the interpretation of landforms. Paleolithologic maps showing [illegible] sediment patterns suggest whether rocks were laid in depths of strong wave action or in quiet water of deeps or broad shoals; lines of equal sediment property, isoliths, can be drawn for many parameters. The organisms in the strata give strong indication of the environments of deposition, the paleoecology. From the study of these and other data, judgments can be reached of water depths and current flows that suggest bathymetry such as is expressed in bathymetric maps. Comparisons between the kinds of rocks and the interpretations of elevations of sources and depths of sites of deposition lead to judgments on regional stability and tectonism. *See* PALEOCLIMATOLOGY; PALEOECOLOGY.

Stratigraphic maps. In addition to studies that concern but instants of time, paleogeography in its broader sense may include the regional distribution of stratigraphic data, such as thicknesses of rocks representing a considerable span of time. The plotting of thicknesses of sedimentary rocks lying between two surfaces of deposition is by isopachs, or lines of equal thickness, on isopach maps. If the surfaces of deposition at the top and at the base of the sequence mapped were horizontal planes, the thickness would represent the amount of warping or deformation of the lower plane prior to the formation of the upper, assuming a stable sea level. As depositional surfaces deviate from horizontality, isopach maps only approach being measures of structural change. The sequence within a time span or stratigraphic interval in an area includes rocks that may differ appreciably both laterally and vertically. Maps showing the ratios of rocks of different kinds, or their constituents, within a stratigraphic interval of some thickness representing a considerable span of time are lithofacies maps, in comparison with paleolithologic maps representing a single surface of deposition. Lithofacies maps can show such ratios as those of sand to silt, or calcium carbonate to calcium magnesium carbonate, or terrigenous sediment to indigenous or precipitated sediment. As long as the average represents conditions that prevailed through the time, lithofacies give evidence of land sources, depths, and other geographic factors. The difficulties of determining precise planes of synchroneity are such that lithofacies maps are the most commonly used expressions of sediment distributions. See FACIES (GEOLOGY).

Tectonic interpretation. A succession of paleogeographic maps showing changes in distributions of the many aspects derived from the study and interpretat...

tec

tre

an effect of eustatic movements. A spread over great areas may involve rise of only a few scores of feet but entail a tremendous volume of marine water. Such changes have been attributed to the addition of water through melting of waning glaciers, some of them may, however, be due to structural changes in ocean basins. On the other hand, the sea can spread because the crust of the earth subsides along the coasts of lands or retreat because the lands rise, such changes in relative elevation of the land through warping movements are

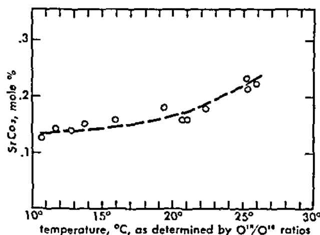


Fig 3. Relationship between strontium content and temperature in articulate brachiopods.

families. The samples include, however, pairs of conspecific and congeneric forms from different temperature regimes. In the echinoid *Dendraster* the relationship between Sr content and temperature differs from that in the articulate brachiopods. The Sr content in the echinoids studied decreases with elevation in environmental temperatures.

Organisms grown under either controlled or natural conditions show that the Sr content of calcareous skeletons is also dependent upon the chemistry of the waters Fresh-water gastropods (*Physa*) grown in waters of increasing Sr content show a corresponding increase of strontium in their shells. The shells of articulate brachiopods taken from the hypersaline waters of the Mediterranean have a higher Sr content than species which grew in the same temperature regime but in mean ocean waters.

Oxygen isotopes. For the relationship of the O^{18}/O^{16} *ratios of the carbonate skeletons to temperature* and to the water chemistry (isotopic methods of geothermometry) *see* GEOLOGIC THERMOMETRY.

Paleoecologic applications. The foregoing demonstrates that certain chemical and mineralogical properties of the skeletons of marine organisms are influenced by the temperature and chemistry of sea water. It should therefore be possible, by studying the skeletal carbonates of marine fossils, to determine the temperature and chemistry of the oceans of the past. Because both environmental factors influence organisms simultaneously, neither should be studied independently. It is essential to determine that the chemistry of the skeletons is unaltered by diagenetic change *See* DIAGENESIS.

These difficulties may be largely overcome by considering, in single fossil specimens, all the properties that are known to be influenced by the two environmental factors. Fossils chosen from a single class have been studied in this manner. The O^{18}/O^{16} ratios, the Mg concentrations, and the Sr concentrations in fossil articulate brachiopods ranging in age from the Pliocene to the Late Mississippian have been determined The temperature values based on these determinations agree within 3C° in each specimen. The results indicate that the shells are chemically unaltered, and that the organisms grew in waters chemically similar to mean ocean water today. If the organisms had lived in insolated water, the temperature values indicated by the O^{18}/O^{16} ratios would be noticeably lower than those indicated by the trace elements A Permian brachiopod which has been studied shows this effect. In samples that had been diagenetically altered by fresh water, the temperature values determined by oxygen isotopes would be higher than those indicated by the trace elements. Several late Paleozoic brachiopod samples appear to have been altered in this way. The close agreement of the temperature determination in the bulk of the samples studied indicates that the chemistry of mean ocean water has been essentially the same for the last 2.5×10^9 years.

Several of the unaltered fossil brachiopods used in the above study are from assemblages containing belemnites which were used for paleotemperature determinations by means of the O^{18}/O^{16} method alone. The close agreement between temperature values determined for the brachiopods and the belemnites adds to one's confidence in paleotemperature determinations, based largely on belemnite rostra, for the Middle and Upper Cretaceous.

Serpulid worm tubes from the Upper Cretaceous of the Coon Creek formation in southern Tennessee illustrate the application of mineralogical studies to paleoecology. The worm tubes consist of an outer calcitic and an inner aragonitic layer Longitudinal cross sections of the tubes show rhythmic variations of the diameters of the calcitic and aragonitic layers, similar to those found in Recent species in response to environmental temperatures. [H.A.L.]

Bibliography: K. E. Chave, Aspects of the biogeochemistry of magnesium: 1, Calcareous marine organisms, *J. Geol.*, 62(3):266–283, 1954; H. A. Lowenstam, Systematics, paleoecologic and evolutionary aspects of skeletal building materials, in Status of invertebrate paleontology, 1953, *Bull. Mus. Comp Zool.*, 112:287–317, 1954; H. A. Lowenstam and S. Epstein, Paleotemperatures of the post-Aptian Cretaceous as determined by the oxygen-isotope method, *J. Geol.*, 62(3):207–248, 1954; H T. Odum, Biogeochemical deposition of strontium, *Inst. Marine Science*, 4(2):38–114, 1957; O. H. Pilkey, Effects of water temperature and salinity on skeletal magnesium and strontium uptake by *Dendraster* (abstract), *Geol. Soc. Am., Rocky Mount. Sect.*, 18–19, 1959.

Paleogeography

The geography of the geologic past; although the term is commonly associated with maps, it concerns all aspects of physical area character in the geologic past that can be determined from the study of rocks.

Geography deals with the face of the earth at a particular time, the present or some interval in the historic past. Paleogeography in the strictest sense concerns the geography of moments or very short time spans in the past. It considers the distribu-

tions of lands and seas, their elevations, depths, and forms. A number of analogous matters commonly are treated as a part of paleogeography, though really stratigraphic, involving spans of time during which a thickness of sediments was laid. See STRATIGRAPHY.

Time and correlations. Paleogeography generally involves the correlation of the rock record and its at the time being considered. In a local area province, the criteria may be lithic, inasmuch conditions may have been such that a single [illegible] of sediment was synchronously deposited throughout a considerable area of land or sea. The identification of time is substantiated by study of organisms, for fossils are the principal means of making time correlations among distant places. [illegible]ope ratios among uranium-lead, thorium-lead, potassium-argon, strontium-rubidium and carbon isotopes are used increasingly in varying degrees of accuracy for differing spans of time. The problems of correlation are in the domain of stratigraphy, but they are basic to paleogeography. See [illegible]CK (AGE DETERMINATION).

Map projections. The preparation of paleogeographic maps involves not only the same problems of distortion of shapes, forms, and areas that are encountered in all map projections, but there are additional difficulties. Most available geologic maps fail to present the proper relative positions of the rocks at the time that the paleogeographic map portrays. When rocks are deformed by folding or faulting after their deposition, the positions and directions between points on opposite sides of each fold or fault have been changed. If a continent has moved relative to another, or a great mountain range has risen through compression, significant deviations develop in the subsequent and present geography from that prior to the deformation. Maps that reconstruct the original relative positions of rocks are known as palinspastic maps. Generally the limits of error in the paleogeography are so great that the present geographic base maps are reasonably satisfactory.

Paleogeographic maps. The simplest forms of paleogeographic maps show the distribution of lands and seas. The study of sedimentary petrology permits determination of the nature of the source lands of the time. Orientations of depositional structures such as ripple marks, cross stratification, flow casts, and elongations of particles and organisms reveal directions of streams, currents, and winds, aspects of paleoclimatology. The elevations of lands can be shown approximately by hypsometric contours of elevation, with colors or shades as used for present geography; more often lands are shown by standard geomorphic landform symbols that give better impressions of the character of lands and less emphasis on elevations. Paleogeologic maps, showing the pattern of rocks on the surface at a past time, aid in the interpretation of landforms. Paleolithologic maps showing bottom sediment patterns suggest whether rocks were laid in depths of strong wave action or in quieter water of deeps or broad shoals; lines of equal sediment property, isoliths, can be drawn for many parameters. The organisms in the strata give strong indication of the environments of deposition, the paleoecology. From the study of these and other data, judgments can be reached of water depths and current flows that suggest bathymetry such as is expressed in bathymetric maps. Comparisons between the kinds of rocks and the interpretations of elevations of sources and depths of sites of deposition lead to judgments on regional stability and tectonism. See PALEOCLIMATOLOGY; PALEOECOLOGY.

Stratigraphic maps. In addition to studies that concern but instants of time, paleogeography in its broader sense may include the regional distribution of stratigraphic data, such as thicknesses of rocks representing a considerable span of time. The plotting of thicknesses of sedimentary rocks lying between two surfaces of deposition is by isopachs, or lines of equal thickness, on isopach maps. If the surfaces of deposition at the top and at the base of the sequence mapped were horizontal planes, the thickness would represent the amount of warping or deformation of the lower plane prior to the formation of the upper, assuming a stable sea level. As depositional surfaces deviate from horizontality, isopach maps only approach being measures of structural change. The sequence within a time span or stratigraphic interval in an area includes rocks that may differ appreciably both laterally and vertically. Maps showing the ratios of rocks of different kinds, or their constituents, within a stratigraphic interval of some thickness representing a considerable span of time are lithofacies maps, in comparison with paleolithologic maps representing a single surface of deposition. Lithofacies maps can show such ratios as those of sand to silt or calcium carbonate to calcium magnesium carbonate, or terrigenous sediment to indigenous or precipitated sediment. As long as the average represents conditions that prevailed through the time, lithofacies give evidence of land sources, depths, and other geographic factors. The difficulties of determining precise planes of synchroneity are such that lithofacies maps are the most commonly used expressions of sediment [illegible]

[illegible] ...ssion of paleo- [illegible] changes in distributions of the many aspects derived from the study and interpretation of the rocks provides a basis for tectonic interpretations. Seas can spread or retreat because of rise or fall of sea level, that is as an effect of eustatic movements. A spread over great areas may involve rise of only a few scores of feet but entail a tremendous volume of marine water. Such changes have been attributed to the addition of water through melting of waning glaciers, some of them may, however, be due to structural changes in ocean basins. On the other hand, the sea can spread because the crust of the earth subsides along the coasts of lands or retreat because the lands rise; such changes in relative elevation of the land through warping movements are

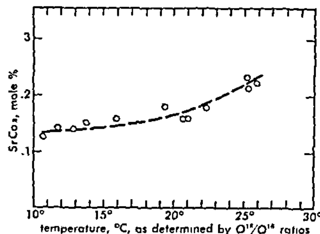


Fig. 3. Relationship between strontium content and temperature in articulate brachiopods.

families. The samples include, however, pairs of conspecific and congeneric forms from different temperature regimes. In the echinoid *Dendraster* the relationship between Sr content and temperature differs from that in the articulate brachiopods. The Sr content in the echinoids studied decreases with elevation in environmental temperatures.

Organisms grown under either controlled or natural conditions show that the Sr content of calcareous skeletons is also dependent upon the chemistry of the waters. Fresh-water gastropods (*Physa*) grown in waters of increasing Sr content show a corresponding increase of strontium in their shells. The shells of articulate brachiopods taken from the hypersaline waters of the Mediterranean have a higher Sr content than species which grew in the same temperature regime but in mean ocean waters.

Oxygen isotopes. For the relationship of the O^{18}/O^{16} ratios of the carbonate skeletons to temperature and to the water chemistry (isotopic methods of geothermometry) see GEOLOGIC THERMOMETRY.

Paleoecologic applications. The foregoing demonstrates that certain chemical and mineralogical properties of the skeletons of marine organisms are influenced by the temperature and chemistry of sea water. It should therefore be possible, by studying the skeletal carbonates of marine fossils, to determine the temperature and chemistry of the oceans of the past. Because both environmental factors influence organisms simultaneously, neither should be studied independently. It is essential to determine that the chemistry of the skeletons is unaltered by diagenetic change. See DIAGENESIS.

These difficulties may be largely overcome by considering, in single fossil specimens, all the properties that are known to be influenced by the two environmental factors. Fossils chosen from a single class have been studied in this manner. The O^{18}/O^{16} ratios, the Mg concentrations, and the Sr concentrations in fossil articulate brachiopods ranging in age from the Pliocene to the Late Mississippian have been determined. The temperature values based on these determinations agree within 3C° in each specimen. The results indicate that the shells are chemically unaltered, and that the organisms grew in waters chemically similar to mean ocean water today. If the organisms had lived in insolated water, the temperature values indicated by the O^{18}/O^{16} ratios would be noticeably lower than those indicated by the trace elements. A Permian brachiopod which has been studied shows this effect. In samples that had been diagenetically altered by fresh water, the temperature values determined by oxygen isotopes would be higher than those indicated by the trace elements. Several late Paleozoic brachiopod samples appear to have been altered in this way. The close agreement of the temperature determination in the bulk of the samples studied indicates that the chemistry of mean ocean water has been essentially the same for the last 2.5×10^8 years.

Several of the unaltered fossil brachiopods used in the above study are from assemblages containing belemnites which were used for paleotemperature determinations by means of the O^{18}/O^{16} method alone. The close agreement between temperature values determined for the brachiopods and the belemnites adds to one's confidence in paleotemperature determinations, based largely on belemnite rostra, for the Middle and Upper Cretaceous.

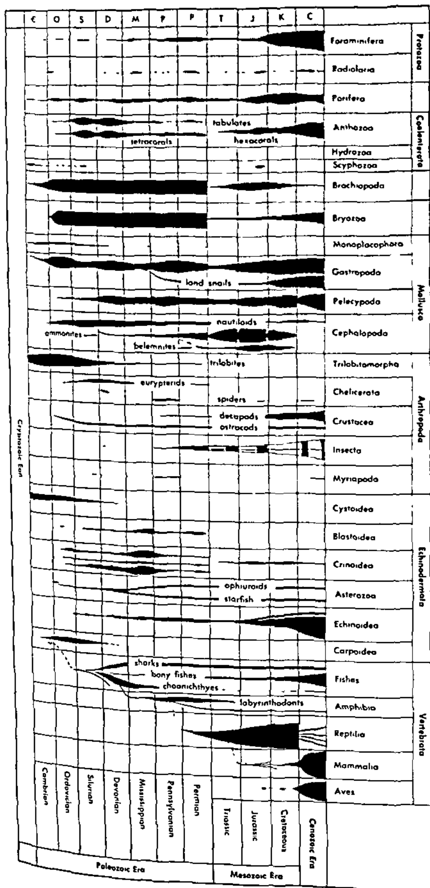
Serpulid worm tubes from the Upper Cretaceous of the Coon Creek formation in southern Tennessee illustrate the application of mineralogical studies to paleoecology. The worm tubes consist of an outer calcitic and an inner aragonitic layer. Longitudinal cross sections of the tubes show rhythmic variations of the diameters of the calcitic and aragonitic layers, similar to those found in Recent species in response to environmental temperatures. [H.A.L.]

Bibliography: K. E. Chave, Aspects of the biogeochemistry of magnesium: 1, Calcareous marine organisms, *J. Geol.*, 62(3):266–283, 1954; H. A. Lowenstam, Systematics, paleoecologic and evolutionary aspects of skeletal building materials, in Status of invertebrate paleontology, 1953, *Bull. Mus. Comp. Zool.*, 112:287–317, 1954; H. A. Lowenstam and S. Epstein, Paleotemperatures of the post-Aptian Cretaceous as determined by the oxygen-isotope method, *J. Geol.*, 62(3):207–248, 1954; H. T. Odum, Biogeochemical deposition of strontium, *Inst. Marine Science*, 4(2):38–114, 1957; O. H. Pilkey, Effects of water temperature and salinity on skeletal magnesium and strontium uptake by *Dendraster* (abstract), *Geol. Soc. Am., Rocky Mount. Sect.*, 18–19, 1959.

Paleogeography

The geography of the geologic past; although the term is commonly associated with maps, it concerns all aspects of physical area character in the geologic past that can be determined from the study of rocks.

Geography deals with the face of the earth at a particular time, the present or some interval in the historic past. Paleogeography in the strictest sense concerns the geography of moments or very short time spans in the [illegible] considers the distribu-



Geologic distribution of animal life

called epeirogenic Eustatic movements cause universal advances and retreats, but epeirogenic movements can be provincial or local. The distribution of sea and land depends on the balance between these two sorts of movements. These are the most general concern of paleogeography. See SEA LEVEL FLUCTUATIONS; WARPING, EARTH CRUST.

Knowledge of the geography of the past has not accumulated to a degree that there are world atlases of paleogeography; in fact, there are very few maps showing world paleogeography, and these are rather simple and crude Series of maps have been prepared for such limited areas as the British Isles, in some detail for single systems in the United States, and in more general form for the paleogeography of North America. Paleogeography represents an end toward which the stratigraphic geologist directs his investigations.

[M.K.]

Paleogeology

The geology of the past, but a term applied particularly to the interpretation of the rocks at a surface of unconformity, that is, an old erosion surface concealed by the deposition of overlying sedimentary rocks. A paleogeologic map showing the distribution pattern of rocks beneath an unconformity can be interpreted like a geologic map of the present surface, permitting recognition of such structural features as anticlines and synclines having older and younger rocks on their axes Such maps suggest the possible sites of petroleum reservoirs beneath concealing sediments, and indicate likely channels of fluid migration. See UNCONFORMITY [M K]

Paleontology

The branch of science dealing with the animal life of the geologic past. Its counterpart, paleobotany, is concerned with ancient plant life. Both are based upon the study of fossils. Paleontology is variously divided according to the materials and the objectives of study. Thus, micropaleontology has to do with fossil organisms too small to study with the unaided eye, and macropaleontology is concerned with the larger fossils The distinction is useful since different techniques of preparation and study are involved, and specialists tend to concentrate in one field or the other. For similar reasons a distinction is commonly made between invertebrate and vertebrate paleontology. A different subdivision is based upon the objectives of study: biologic paleontology is concerned with the morphology, systematics, and life habits of the ancient organisms; stratigraphic paleontology is concerned with the use of fossils in dating and correlating the stratified rocks. In a broad sense paleontology is concerned with the evolution and geologic history of life on earth. See FOSSIL; GEOLOGY; MICROPALEONTOLOGY; PALEOBOTANY; STRATIGRAPHY.

Paleontology and evolution. Fossils provide a record of life as it existed from age to age and the documentary evidence for evolution The discovery of a series of fossil horses in the Cenozoic formations of western North America showing progressive change from a three-toed to a one-toed condition was one of the first convincing lines of evidence that life has evolved. Since then, other sequences of fossil organisms have been discovered These also indicate progressive change with time. Some of these show small-scale changes from one genus to another, some link one family to another, and yet others provide connecting links between larger taxa For example, the first bird, *Archeopteryx*, which had a long tail and clawed wings, is an almost ideal link between primitive reptiles and birds. The theriodont reptiles of Permian and Triassic time likewise show many gradual changes in the structure of skull, teeth, and limbs leading to the mammals, so that the exact stage at which mammals began is difficult to decide.

Because deposition of sediments with their entombed fossils is commonly discontinuous at any given locality, the evolution is incompletely recorded, and fossil species commonly appear distinct and well defined; but in places where the record is more complete, it is difficult to draw specific boundaries. In the Jurassic shales about Petersboro, England, for example, Brinckmann collected abundant ammonites from 1300 thin zones and, by applying biostatistics, proved a finely graded series that linked five previously described species of the genus *Zugocosmoceras* and five of the genus *Cosmoceras* into completely intergrading sequences in which specific limits disappeared. See EVOLUTION, ORGANIC.

Paleontology and paleogeography. The distribution of fossils, together with the lithologic character of deposits, provides the basis for determining the past distribution of lands and seas. Fossil marine organisms record the spread of inland seas that have now vanished, even in regions that have since been uplifted into mountains. Identical or closely similar fossil faunas in land masses now separated by seas also record land bridges that have since disappeared. The large mammals of the East Indies, for example, obviously came from the mainland of Asia when the Sunda Shelf was emergent The sudden arrival of the mastodons in North America in Miocene time likewise indicates a land bridge across the Bering Strait On the contrary, the quite independent evolution of the South and North American mammals during most of Cenozoic time proves that these continents were then isolated; and the great migration of several orders of South American mammals into North America, and vice versa, about the end of Pliocene time clearly dates the origin of the Isthmus of Panama See PALEOGEOGRAPHY

Paleontology and paleoecology. Fossils also throw much light on the environment under which the sedimentary rocks of past ages accumulated Miocene and Pliocene floras in Nevada are closely similar to, and in part identical with, those now growing in the low, humid Gulf Coastal Plain. It is therefore evident that at that time the region was lower and more humid than now. Assemblages of fossils in ancient marine formations may likewise show whether the sea floor was firm or soft, clear

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	*Mississippian*	*Pennsylvanian*	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

designated as systems, is based on early geologic studies in Europe *See* STRATIGRAPHIC NOMENCLATURE.

Throughout most of the world, the Paleozoic Era is considered to embrace six geologic periods that correspond to and are designated by the names of geologic systems In order from oldest to youngest, with estimate of their respective duration in millions of years given in parentheses, these are Cambrian (80), Ordovician (80), Silurian (40), Devonian (55), Carboniferous (55), and Permian (25). Each of these is treated by a separate article. *See* GEOLOGY

[R C M.]

Palladium

Chemical element number 46, pallad... white and ... num in ma... and abundance.



USES The most important use of pure palladium is in low-current electrical contacts, especially in telephone equipment Palladium supported on carbon or alumina is used as a catalyst for hydrogenation and dehydrogenation in liquid- and gas-phase reactions Palladium on alumina is particularly effective in permitting traces of oxygen in hydrogen to combine to form water which is easily removed

Palladium alloyed with 4 5% ruthenium has the proper hardness for jewelry. Palladium is unique in the number of metals with which it can be alloyed and generally produces ductile solid solutions. When alloyed with other precious metals and base metals, it is used for dental purposes, electrical contacts, and special resistance wires Alloys of palladium, silver, and copper, with and without added manganese, are used as strong brazing alloys useful at quite high temperatures. *See* ALLOY; ALLOY STRUCTURES.

Metallurgical extraction. In one type of refining, a filtrate containing palladium(II) chloride is formed This is treated with ammonium hydroxide and heated to form a solution of tetrammine palladium chloride Dichlorodiammine palladium is then precipitated with hydrochloric acid. This is quite soluble in cold dilute ammonia. Palladium may also be extracted by virtue of its ability to form insoluble cyanide, iodide, and dimethylglyoxime derivatives. The separation of palladium, platinum, and rhodium from the other platinum metals may be made by fusing with lead. Ruthenium, osmium, and iridium are insoluble in lead except at very high temperatures, whereas the other platinum metals as well as gold and silver dissolve.

Physical and chemical properties. Palladium is a soft, very ductile metal and may be worked down cold to thin foil and fine wire. Because of its softness, it is often alloyed with other precious or base metals to increase its hardness and tensile strength.

Properties of palladium

Property	Value
Atomic weight	106 7
Crystal structure	face-centered cubic $a = 3.89$ A at 20°C
Density (at 20°C)	12 00 g/cm³
Melting point	1552°C
Linear thermal expansion coefficient (per °C at 20°C)	12.4×10^{-6}
Specific heat (at 0°C)	0 0584 cal/°C
Thermal conductivity (at 20°C)	0 168 cal/(cm²)(°C)(sec)
Electrical resistivity (at 20°C)	10 8 μohm-cm
Temperature coefficient of electrical resistivity (per °C, 0–100°C)	0 0038
Modulus of elasticity	16×10^6 psi
Tensile strength—hard	55,000 psi
Tensile strength—annealed	25,000 psi

A thin oxide surface film is formed on palladium when it is heated in air to 360–800°C, but this decomposes at higher temperatures to leave the metal bright Hydrogen is readily absorbed by palladium and will diffuse through heated palladium at a relatively rapid rate This property is used in hydrogen purifiers which pass hydrogen but no other gas

Palladium in ordinary atmospheres is resistant to tarnish but will tarnish slightly in outdoor exposures in sulfur-contaminated atmospheres At room temperature, palladium is resistant to hydrofluoric, phosphoric, perchloric, and acetic acids Hydrochloric and sulfuric acids attack palladium slightly; the metal is readily attacked by nitric acid, ferric chloride, and by moist chlorine, bromine, and iodine

Palladium compounds. Palladium has two principal oxidation states, 2+ and 4+, the more common being 2+ Palladium chlorides and related compounds are the m...

or muddy, and whether the bottom water was well aerated or stagnant and foul. They may also indicate whether the water was warm or cold. Indeed, paleotemperature study based on the O^{18}/O^{16} ratios in fossil shells may record the absolute temperature at the time of deposition. *See* GEOLOGIC THERMOMETRY; PALEOECOLOGY.

Cryptozoic history of life. For all practical purposes, the recorded history of life began at the base of the Cambrian System in rocks deposited approximately 500,000,000 years ago. Before this was the Cryptozoic Eon with rocks ranging up to nearly 3,000,000,000 years old, in which animal fossils are virtually lacking and plants are limited to rare and extremely primitive types. In the Gunflint chert of the Huronian System north of Lake Superior microscopic filaments of blue-green algae and fungi have been found. The enclosing rocks are dated by radioactive isotopes as being about 1,500,000,000 years old Limy deposits, even reefs, formed by microscopic algae, are more widespread in both the Huronian and Beltian rocks, but no higher types of plants are known

The only certain evidence of animal life in the Cryptozoic Eon consists of trails and burrows of wormlike creatures in the Siyeh shale of the Beltian System, known to be 1,000,000,000 years old. Two considerations indicate that animal life was probably abundant in the seas and even highly diversified for perhaps 1,000,000,000 years before the Cambrian: (1) the great diversity of the earliest Cambrian faunas implies a long antecedent evolution, and (2) a vast amount of carbon embedded in the Precambrian sedimentary rocks is distributed in the same manner as it is in later formations where it is known to be the residue of organic matter buried with the sediments

Insight as to the probable nature of this early life may be drawn from the unique fauna of the Burgess shale of Middle Cambrian age near Field, British Columbia. In this black, slaty shale an array of soft-bodied animals is preserved in the form of delicate films of carbon on the bedding planes. Many of these show the outline of the bodies with delicate appendages, and some even show the viscera In this fauna of 70 genera and 130 species, the great majority of the species were soft-bodied and could not have been preserved in any other form. The fauna includes sponges, jellyfish, probable holothurians, an onychophoran, and a great variety of primitive arthropods. Although some of these range up to several inches in length, not one of the soft-bodied forms has ever been found elsewhere; only the exceptional local conditions permitted them to be preserved. However, the abundance and variety of animals in this locality make it appear certain that such animals were abundant in the mid-Cambrian seas. It is altogether probable that a comparable diversity of soft-bodied creatures existed in the Precambrian seas but was not preserved. It was not until animals had developed armor in the form of chitinous tests or shells that they began to leave a significant fossil record The lack of armor in previous ages cannot be attributed to lack of sufficient lime in the Precambrian seas, as Daly supposed, since the predominance of chitinous and phosphatic shells in the Lower Cambrian formations proves that calcium carbonate was not required. It appears probable that the development of firm protective armor followed rapidly upon the evolution of predatory habits and that this occurred about the beginning of Cambrian time.

History of Phanerozoic invertebrates. The Phanerozoic Eon includes the time since the beginning of the Cambrian Period. The accompanying figure shows in a very simplified form the history of invertebrate animal life as known from the fossil record. It may be noted that most modern phyla that bear tests or shells were present in Cambrian time, that all others had appeared by the middle of the Ordovician Period, and that they were already clearly differentiated when the fossil record begins. On the other hand, a number of the classes and orders did not appear until Ordovician time and some much later. It is notable also that a great and progressive evolution continued until the end of the Paleozoic Era, when several groups that had been dominant became extinct and other groups subsequently expanded to take their place. Thus the faunal break at the end of the Paleozoic Era is greater than any other in the geologic record A second major break occurred at the end of the Mesozoic Era, when the ammonites and belemnites disappeared and most of the reptiles that had dominated the lands during the era died out.

[C.O D]

Bibliography: E. H Colbert, *Evolution of the Vertebrates*, 1955; C O. Dunbar, *Historical Geology*, 2d ed , 1960, R. C. Moore, *Introduction to Historical Geology*, 2d ed , 1958; R. C. Moore, C. G. Lalicker, and A G. Fischer, *Invertebrate Fossils*, 1952; A S Romer, *Vertebrate Paleontology*, 2d ed , 1945, R. R. Shrock and W. H Twenhofel, *Principles of Invertebrate Paleontology*, 2d ed , 1953; R. A Stirton, *Time, Life, and Man*, 1959; H. H Swinnerton, *Outlines of Palaeontology*, 3d ed , 1947.

Paleozoic

One of the major divisions of earth history, with time span of approximately 335,000,000 years according to radioactive age measurements. The Paleozoic Era (signifying ancient life era) comprises the earlier two-thirds of the so-called Phanerozoic Eon (evident life), which is characterized by relatively abundant records of the plant and animal life of the earth's past preserved as fossils in successive layers of sedimentary rocks. The deposits of Paleozoic age contrast very strongly with older rocks, collectively termed Cryptozoic (nonevident life), not only in richness of organic content but in their prevailing lack of metamorphic alteration and generally much more simple structure. Paleozoic formations are well represented by wide distribution and great aggregate thickness on all of the continents, but classification of their main divisions,

Palpigradi

An order of rare arachnids comprising 21 known species from tropical and warm temperate regions. American species occur in Texas and California. All are minute, whitish, eyeless animals, varying from 0.68 to 2.8 mm in length, that live under stones, in caves, and in other moist, dark places. The elongate body terminates in a slender, multisegmented flagellum set with setae. In a curious reversal of function, the pedipalps, the second pair of head appendages, serve as walking legs. The first pair of true legs, longer than the others and set with sensory setae, has been converted to tactile appendages which are vibrated constantly to test the substratum. See ARACHNIDA [W.J.GH.]

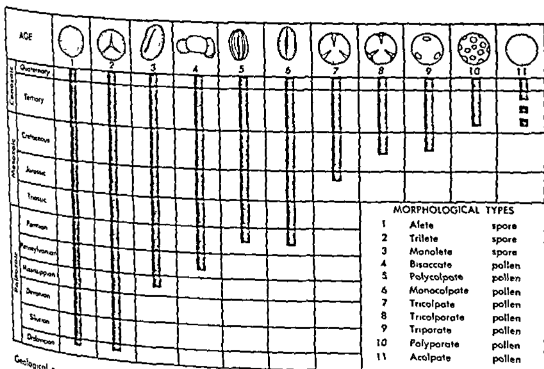
Palynology

The study of spores and pollen. The discussion here is primarily confined to the occurrence and study of fossil spores and pollen, a branch of micropaleontology. This study began in the early years of this century, but the extensive use of spore and pollen analyses in stratigraphic studies did not begin until about 1930. Since then spores and pollen grains have become as important fossil indicators as the Foraminifera and ostracods. See MICROPALEONTOLOGY.

Morphology. Spores, except those of fungi, are one-celled reproductive bodies that give rise to the gametophyte generation of plants. Pollen grains are microgametophytes, consisting of two to several cells, all of which are contained within the walls of the microspores from which they developed. Spores in the fossil record are from the nonflowering members of the plant kingdom, whereas pollen grain production is confined to the gymnosperms and angiosperms. The general features which distinguish spores from pollen are the mechanisms by which each germinates. Spores, except those of fungi, usually possess either single (monolete) or Y-shaped (trilete) sutures. Some spores are without any germinal openings (alete), and some of the fungal spores possess a simple pore. Pollen grains may have one or many germinal furrows (colpate grains). The furrows are simple, or they may contain a pore in the median longitudinal position (colporate grains). Other pollen grains may have one or many pores (porate grains). The number and distribution of the colpi and pori constitute the basis for most classifications used in distinguishing fossil pollen. Other characters utilized in the description and classification of both spores and pollen are shape, size, ornamentation, and special morphological features such as air bladders and elaters.

Fossil record. The paleontologic record of spores begins in the Cambrian rocks, but fossil spores are not abundant until Devonian time. Pollen fossils of primitive gymnosperms are present in the Devonian rocks, but are not abundant until Mississippian time. Pollen grains of angiosperms are abundant in Lower Cretaceous rocks, and there is evidence that they were present in the Jurassic period. The accompanying diagram shows the geologic range of spore and pollen types.

The chemical nature of the spore and pollen walls makes them resistant to decay and consequently they are preserved in most sedimentary environments and rocks. Other factors which add to



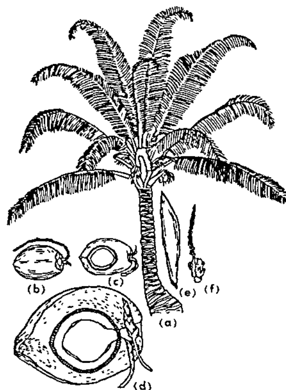
Geological ranges of spore and pollen types

processes. Palladium nitrate and nitrite compounds are also useful for much the same purposes.

Palladium monoxide, PdO, and dihydroxide, $Pd(OH)_2$, are used as sources of palladium catalysts. For a discussion of the natural occurrence of palladium, *see* PLATINUM. [H.J.A.]

Palm

Shrubs, trees, or vines of the family Palmae. The stems are generally unbranched and, in structure, more nearly comparable to cornstalks than to true trees. There are at least 1500 species, perhaps many more. A few are temperate zone plants, but most are tropical species. Various palms are much used

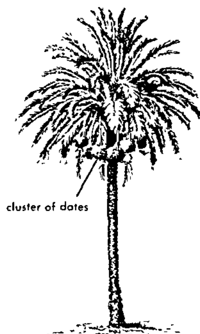


Common coconut palm (*Cocos nucifera*) (*a*) Plant (*b*) Coconut fruit entire (*c*) Longitudinal section of fruit when ripe. (*d*) Longitudinal section to show growing seedling with spongy foot filling cavity. (*e*) Spathe enclosing flower clusters. (*f*) Single cluster showing staminate flowers above and pistillate flower below. (*From O. Degener, Ferns and Flowering Plants of Hawaii National Park, 1930*)

as ornamentals. For the inhabitants of tropical regions, the palms rank with the grasses in economic importance, providing food, shelter, clothing, and numerous other necessities. Economically, the coconut palm, *Cocos nucifera*, holds first place, and the date palm, *Phoenix dactylifera*, is second in importance. It is estimated that there are more than 1000 uses of palms and their products. *See* MONOCOTYLEDONEAE; PALMALES. [P.D.S.]

Palmales

An order of the plant subclass Monocotyledoneae with a single family Palmae including 200 genera

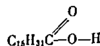


Common date palm (*Phoenix dactylifera*). (*From E. Le Maout and J Decaisne, Traité Général de Botanique Descriptive et Analytique, Librairie de Firmin-Didot et Cie*)

and probably more than 1500 species. The taxonomy of the group is exceedingly difficult and not completely known. All are woody plants—shrubs, vines or trees Leaves are large, usually in a terminal crown, simple, and palmately or pinnately compound Distribution is mainly in the tropics, and for the people in this region, the palms are second only to the grasses in economic importance. They are a source of starch or sugar and yield valuable fruits (coconuts and dates). They supply material used in construction, textile fibers, and thatch for roofing The leaves are made into mats, baskets, and hats Waxes (carnauba wax) and oils are obtained from some of the palms. *See* CARNAUBA; COCONUT; DATE, PALM; VEGETABLE IVORY; *see also* EMBRYOPHYTA; MONOCOTYLEDONEAE; PLANT KINGDOM. [P D S]

Palmitate

A salt (soap) or ester of palmitic acid



in which the acidic hydrogen has been replaced by a metal or an organic radical. Palmitates occur in nature, chiefly as the glyceryl esters, found in substantial amounts in animal and vegetable fats The esters of long-chain alcohols are known as waxes. Simple esters have limited uses Alkali metal palmitates are soluble in water, and with the similar stearates and oleates, are the major components of toilet and laundry soaps. Other metal soaps are used in lubrication greases, pharmaceuticals, and cosmetics, and as waterproofing agents and fungicides. *See* CARBOXYLIC ACID; ESTER; FAT AND OIL, NONEDIBLE; SOAP AND DETERGENT. [F H H.]

indicate coastal, upland, or mountainous ter- . Mixtures of several spore and pollen floras are erable in modern estuary deposits and the comable occurrences in older deposits may be exned by similar river transport

etroleum exploration. In the modern search for roleum, palynology has become an important By it, many age and facies correlation probs are being resolved. The search for source and ervoir rocks must utilize paleoecological and ogeographical information secured from many rces, among which spore and pollen studies important. In many rock sections larger fossils scarce or absent. In others the larger fossils are ken by the drilling tools and consequently are of le correlation value. Spores and pollen are freently abundant in these rocks and because they rarely more than 150 μ in diameter, they are t injured by drilling activity. The source beds most petroleum deposits are associated with allow marine and brackish water sediments. In se sediments spores and pollen occur abunntly in association with microscopic marine fosls. They decrease in number and species with stance from shore. When drilling operations are nfined to the exploration of a certain stratum at is presumed to be a source bed, samples are covered from it at several locations and analyzed r both marine and continental microfossils. The esults of the analyses are plotted on a map to show e ratio between continental and marine facies. he larger continental ratios should occur on the ide toward the ancient shoreline, and in the direcion of a possible stratigraphic wedge out, where petroleum may be accumulated.

Stratigraphic correlation. Stratigraphic correlations by spores and pollen over distances in excess of several hundred miles impose problems also inherent with other fossils. These are (1) that environments suitable for preserving fossils often are not continuous, (2) that the floral assemblages may have geographic controls, (3) that floras are sensitive to climatic changes and their composition was often in a state of flux, and (4) that there may be persistence of floristic assemblages in relict communities. Regardless of these difficulties, it is possible to recognize assemblages of each geological period wherever they occur. The recognition of strata deposited in shorter periods of geological time requires intensive study of intermediate localities. Some of the problems of distant correlation are illustrated in the peat and silts of postglacial time. At present there are belts of forest types spread latitudinally across the United States from the East Coast to the Great Plains. During the glacial advances these forests retreated southward and returned northward with the melting of the ice. The sequence of forest succession followed roughly the following order: spruce forest, pine forest, and mixed hardwood forests. Each glacial advance covered different areas from the preceding; consequently, the forest migrations were not similar. Some forest types remained longer in certain areas than in others and the length of postglacial time is variable in different areas. Fluctuations in the pollen spectra of deposits outside of the glacial area, and on the several Wisconsin drifts, may indicate glacial advances and retreats. In order to correlate postglacial deposits of various ages, it is necessary to compare the total pollen succession and to take into account the geography of the deposits. Normally the topmost pollen assemblages of Recent deposits will be correlative in time though they consist, for example, of spruce and pine in northern Ontario, and oak and hickory in southern Ohio. The bottommost pollen assemblages in these areas are not correlative in time even though they are similar, consisting of spruce and pine. Both bottom and top assemblages represent stages in floral succession in which age and climate are factors. Identification of specific interglacial deposits by pollen spectra is not conclusive at present. The same palynological problems exist for the several interglacial intervals

[illegible]

cene, the differentiation of the interglacial deposits must be based on assemblages. *See* POSTGLACIAL VEGETATION AND CLIMATE; *see also* PALEOCLIMATOLOGY; PALEOGEOGRAPHY; PLANT GEOGRAPHY.

[L.R.W.]

Bibliography: G. Erdtman, *An Introduction to Pollen Analysis*, 1943; D. J. Jones, *Introduction to Microfossils*, 1956; H. W. Matthes, *Einführung in die Mikropaläontologie*, 1956; L. R. Wilson, The correlation of sedimentary rocks by fossil spores and pollen, *J. Sediment. Petrol.*, 16(3):110–120, 1946.

Pancarida

A superorder of the subclass Malacostraca. It is comprised of crustaceans of small size, measuring 1–3 or 4 mm, which carry their eggs and embryos in a brood pouch, or marsupium, that is dorsally located and formed by the carapace. The maxillipeds usually have two endites, either an exopodite

[illegible]

seventh and eighth thoracopods.

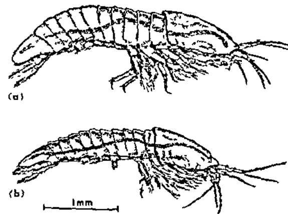


Fig. 1. *Thermosbaena mirabilis* Monod. (a) Female. (b) Male. (After A. Bruun)

the importance of spores and pollen in paleontological studies are that (1) they are minute and are produced in exceedingly large numbers; (2) they are spread widely by wind and water; (3) they possess structures which make them identifiable into groups and species; (4) they are specifically restricted within short geological time divisions; (5) they are useful indicators of paleoecology; (6) they occur in continental as well as in marine deposits and thus make possible the correlation of continental and marine sediments; (7) they are indicators of paleogeographic conditions; and (8) they can be treated statistically to reveal the degree of correlation.

Fossil spores and pollen occur in many types of rock. They are most abundant in organic sediments such as peat and coal. They are absent or rare only in the coarse clastic sediments, in some of the chemically precipitated rocks, and in rocks or soils that have been oxidized before or after lithification.

Recovery and identification. Palynological work requires special techniques of sampling, processing, and study The nature of the investigation determines the sampling method. Channel samples are usually collected from outcrops and well cores. Rock chips or cuttings from wells may be used but the final interpretations should take into account the inclusion of contaminating fossils from higher levels. The usual procedure is to work from the top of the well toward the bottom and to record the first occurrences of index fossils. For peat and soil studies special boring devices are employed in the recovery of samples. *See* INDEX FOSSIL.

The methods used in processing are determined by the sample's lithology. Since most spores and pollen are resistant to the action of acid, these chemicals are used to release the fossils from their matrices. Calcareous sediments are removed by the use of hydrochloric acid and siliceous sediments by hydrofluoric acid. Carbonaceous rocks are given additional treatment with saturated potassium chlorate solution combined with concentrated nitric acid. Potassium bichromate dissolved in sulfuric acid also may be used to process carbonaceous rocks. After digestion or maceration of the rock has been completed, the residues are washed with water and treated with a basic solution, ammonium or potassium hydroxide. Fossil-bearing rocks differ greatly in lithology and each must be treated as a special problem. While being processed, the samples may be subjected to ultrasonic treatment to aid in releasing fine particles of mineral matter adhering to the fossils and for the destruction of aggregates and gels

When the fossils have been concentrated, they may be stored indefinitely in alcohol or other preservatives. Microscope slides are made by placing a drop of the fossil concentrate in warm glycerin jelly that has been spread on a coverslip. The fossils are then gently mixed with the glycerin jelly until spread uniformly. A margin of glass should be left completely around the coverslip. These mounts are dehydrated in an oven and mounted, preparation down, on a microslide using Canada balsam other permanent seal.

The compound microscope used to study fos spores and pollen should have, in addition to t usual optics, an oil immersion objective and pha equipment. A mechanical stage is also necessai In the study of microfossils it is expedient to phot graph the specimens for comparison and recor Cameras using 35-mm film are satisfactory for th purpose.

The identification of near-Recent fossils is a complished by comparison with prepared mode spores and pollen. These fossils occurring in Te tiary and older rocks are identified from the d scriptive literature and typical specimens in co lections. There are several existing philosophies spore and pollen classification and nomenclatur These are based upon the structure and number germinal apertures, ornamentation, and shape the fossils. Owing to the youthful age of the sc ence, universal agreement in these matters has n been attained.

Measurement and correlation. The analyst of microfossil assemblage must determine the numbe of species present in a deposit and their relati abundance. This is done by studying a sample unt only an occasional new form is observed. That num ber varies with the preservation of the fossils an richness of the fossil flora. Fossil assemblages con taining fewer than 25 species usually necessitate count of 200 to 500 fossils per sample for correla tion. Assemblages with more than 25 species ofte require the counting of over a thousand fossils for satisfactory analysis. Several methods of graphi representation are used to illustrate the floristi content, succession, and correlation of the samples These consist of species charts, bar and line graphs and are referred to in literature as the spectra o spore and pollen profiles.

An examination of the graphs and charts result ing from the microscopic study reveals informatior relative to the deposits' age, paleoecology, and paleogeography. Age determination is based upon a knowledge of stratigraphic ranges of the groups, genera, and species. Closest age determination is accomplished at the species and species-assemblage levels. The paleoecology of a fossil assemblage is related to the ancient geographic, climatic, and primary and secondary floral successional controls. These factors are often recognizable in Cenozoic and upper Mesozoic deposits by comparison of the fossil assemblages with similar Recent assemblages The method requires a detailed study of numerous samples through vertical sections to determine the related occurrence and abundance of the fossils. When several adjacent sections are compared, the ecological relations become useful stratigraphic criteria.

Paleoclimatic conditions are roughly determined by recognition of typical subtropical, temperate, or polar floras especially if the spores and pollen grains are associated with marine sediments. Assemblages that are known to represent continental deposits may also reflect the topographic relations

arising from the base of the branch corresponding to the Peracarids, a short distance from the bifurcation leading to the Eucarida *See* MALACOSTRACA. [T.M.O.]

Pancreas

A composite gland in most vertebrates, containing both exocrine cells which produce and secrete enzymes involved in digestion, and endocrine cells, arranged in separate islets which elaborate at least two distinct hormones, insulin and glucagon, both of which play a role in the regulation of metabolism and particularly of carbohydrate metabolism. *See* CARBOHYDRATE METABOLISM.

EMBRYOLOGY

Ontogenetically, the organ is of entodermal origin. It arises initially from a series of outpocketings of the embryonic digestive tract, which eventually fuse, to various extents in various species, to form a single organ The adult gland in mammals results from the fusion of two such evaginations, one dorsal, opposite the hepatic diverticulum, and one ventrolateral, at the base of the biliary duct In form [illegible]

[illegible] elasmobranch pancreas is derived solely from a dorsal evagination, the ventral components never arising.

The primordia. The pancreatic diverticula make their appearance early in development, the dorsal element preceding that of the ventral component or components In the chick, the pancreatic primordia appear between the third and fourth days of incubation the dorsal primordia arising about the third day and the ventral on the fourth day. The dorsal component in the pig appears at the 4-mm stage, followed by the formation of the ventral diverticulum at the 5-mm stage The pancreatic primordia of amphibians appear between the 8- and 10-mm stages, depending on the species.

Pancreatic ducts and tubules. The base of the evaginations eventually develop into the pancreatic ducts, Santorini's duct from the dorsal rudiment, Wirsung's duct from the ventral Santorini's duct, when it persists, opens directly into the duodenum, while the duct of Wirsung joins the common bile duct. Although both ducts may be retained throughout adult life in some species, such as the chick, horse, and dog one or the other duct usually disappears, leaving a single pancreatic duct Thus Wirsung's duct is the adult pancreatic duct in man sheep ganoid fish, teleost fish, and the frog, while the duct of Santorini serves the adult elasmobranch, pig, and ox

The original pancreatic primordia proliferate into masses of undifferentiated cells which then grow into solid cords of cells These cords eventually are transformed into primitive pancreatic tubules. By a process of budding, the exocrine acini are formed from the primitive tubules. In addition, masses of loosely connected cells, centroacinar cells, are derived from the primitive tubules, and form an anastomosing network between the primary tubule and acinar system.

Endocrine elements. The endocrine cells of the pancreas are grouped in aggregates or islets of varying size which have no connection to the duct system. These aggregates, the islets of Langerhans, are highly vascularized and distributed throughout the organ. The distribution of islets is not uniform, and it has been suggested that this unequal distribution is a reflection of the fact that most of the endocrine cells are derived from the dorsal pancreas This contention is best supported by the evidence presented from studies of certain urodele species in which fusion of the pancreatic diverticula occurs late and is restricted to the formation of a narrow isthmus of tissue between the pancreatic lobes. The existence of a separate giant islet of endocrine cells in certain teleosts also suggests the predominant role of one diverticulum in the development of the endocrine component.

Islet morphogenesis The early endocrine elements of the pancreas arise from the cells of the diverticulum itself and from the primitive tubules These cells form the primary islets of the pancreas, sometimes referred to as the islets of Laguesse Considerable controversy has been waged over the eventual fate of the primary islets The alternative views are (1) that the primary islets degenerate and are replaced by the definitive islet tissue, and (2) that the early islets persist into adult life A second generation of islet cells occurs in the developing pancreas, although many investigators do not make a sharp distinction between primary and secondary islets These secondary islets are derived largely from centroacinar cells and from the cells of the developing duct system It has also been suggested that differentiated acinar cells can, by a process of dedifferentiation, give rise to the secondary islets This contention, however, has been hotly disputed Transitional stages in the formation of these islets of endocrine cells from ductules have been described in the embryonic organ In adult organs of certain primitive species, such as the Elasmobranchii, these transitional stages may persist throughout adult life For example, in some species of elasmobranchs, the islet cells are arranged in sheets next to the duct cells, in others, the islets may be connected to a ductule by a solid cord of cells Phylogenetically, a progression of pancreatic types exists which resembles the development stages occurring in any one of the higher vertebrates

Lower chordates Scattered cells resembling pancreatic cells have been described in the digestive tube of the lancet *Amphioxus* In the lamprey, *Petromyzon*, the pancreas remains embedded in the wall of the intestine, having lost, secondarily, its pancreatic duct The suggestion has been made that in the Cyclostomata like the lamprey the pancreas is primarily endocrine in function Certain anatomical and histological features of the development of the digestive tract of the ammocoete larvae of the lamprey have led several investigators to speculate that this organism represents a transi-

The group includes a unique order, Thermosbaenacea Monod, 1927, with two families, Thermosbaenidae and Monodellidae Taramelli 1954. In the first, there is one genus and one species, *Thermosbaena mirabilis* Monod, 1924; the second also contains one genus, *Monodella* Ruffo, 1949, but three species, *M. stygicola* Ruffo, 1949, *M argentarii* Stella, 1951, and *M halophila* Karaman, 1953.

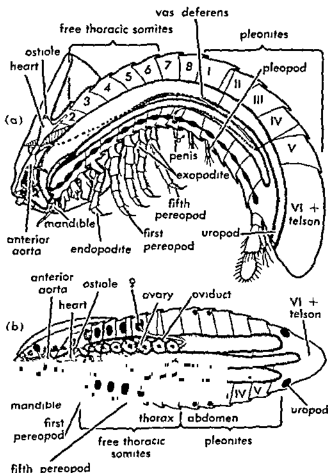


Fig. 2. *Thermosbaena mirabilis* Monod (a) Male (b) Female (After R Siewing)

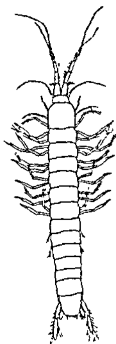


Fig 3 *Monodella halophila* Karaman, male (After S. Karaman)

Ecology. The habitat of the various species is remarkable. *T mirabilis* has been collected from the Roman bath, El Hamma, near Gabès, Tunisia The temperature of the water is 45–48°C. *M stygicola* is known from a small, brackish, subterranean lake near Castromarina, Terra d'Otranto, Italy; *M. argentarii* from a small subterranean lake of Monte Argentaria in the Tyrrhenian mountain range north of Rome; and *M. halophila* from a subterranean pool at Gruž (Dubrovnik), Yugoslavia, and a small subterranean lake near Cavtat, Yugoslavia. All these localities are found in the proximity of the sea, in the transitional zone between the marine and fresh-water environments

Morphology. The body is cylindrical, eruciform, and lacks an external division between the thorax and pleon. Eyes are absent and the cephalon is united with the first thoracomere; the carapace covers the first to third thoracomeres. The telson is united to the last pleonite in *Thermosbaena* and is not united in *Monodella*. The antennule is biramous and the antenna has an exopodite. There are 5 pairs of biramous pereiopods on thoracomeres 2–6 while thoracomeres 7 and 8 are without appendages in *Thermosbaena*. Seven pairs occur in *Monodella* on thoracomeres 2–8. The pereiopods lack epipodites Two pairs of pleonodes occur on the pleomeres I and II, the first pair being jointed. The biramous uropods have a one-jointed endopodite, while the exopodite is two-jointed. The heart is anterior, compact, sacciform, and has two ostioles Oostegites are lacking.

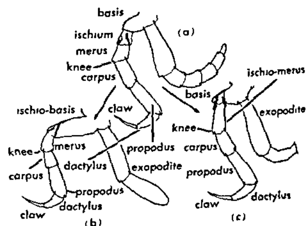


Fig 4. Hypothetical scheme of the evolution of the pereiopods from an ancestral type (a) Ancestral pereiopod (b) First pereiopod of *T mirabilis* (c) Pereiopod 2–5 of *T. mirabilis*. (After R. Siewing)

The Thermosbaenacea have been linked to the Peracarida by T. Monod, to the Syncarida by G. O. Sars, and held to be more or less intermediate between these two groups by E. Stella and Taramelli. In 1958 R. Siewing established, through detailed anatomical studies, that the group must comprise an autonomous division, the Pancarida,

ancreas, has been localized in the anterior portion of the intestine The hypothesis is further strengthened by the observation that in the larva of an Australian lamprey, *Geotria australis*, the intestine grows anteriorly as a pair of blind pouches. These pouches are lined with the granular cell type. It is difficult to escape the temptation to speculate that the ammocoete larva represents an evolutionary stage in which the endocrine pancreas has already separated from its original location and that the ammogeulike cells are preparing to follow suit, namely, the condition in the larva of *G. australis* *The next phylogenetic advance is observed in the* elasmobranch. In these forms, a single dorsal pancreatic primordium and, consequently, the single duct of Santorini develop. Thus, phylogenetically as well as ontogenetically, the dorsal pancreas arises first. The first orders in which a pancreas similar to the mammalian organ appears are the ganoids and teleosts.

Another observation bespeaking the embryonic origin of the pancreas from the epithelium of the primitive digestive tract is the demonstration that cells with staining properties similar to those of the alpha cells *occur* in the pyloric portion of the bovine stomach. Extracts of this region give a positive test for the hormone glucagon

...l types. Within an islet of Langerhans, at least ... distinct *granular* cell types can be distin... ed on the basis of tinctorial differences among ...ranules demonstrable with polychrome stains. ...nt evidence strongly suggests that the acido... alpha cell is concerned with the synthesis of ...ormone glucagon *and* that the basophilic beta ...is the site of insulin synthesis *The third cell* the delta cell, may be a transitional form. In ...ion, a fourth, agranular, *cell* type may be ...ent in some species *Although*, in most species ...islet contains all of *the* representative cell ..., it has been reported that, in the domestic ..., the individual islets *may* be predominantly ...a-cell or beta cell islets. *Considerable* variation ...ts as to the time and order of appearance dur... development of the cell types among *the* various ...cies. In the rat, the beta cell can first be dis...ed in the 18½-day fetus; *alpha* cells cannot ...demonstrated until two days *postpartum* The ...ha cell, on the other hand, is the *first islet* cell ...e recognized in the chick pancreas, appearing ...the eighth day of incubation. *On the twelfth* ..., concomitantly with the appearance of large ...mbers of degenerating cells, beta cells *can* be ...stinguished As yet, no definitive evidence is avail...le which would permit a correlation between the ...ection of a particular cell type and the *produc*...on of the hormonal product of that cell *type How*...ver, on the basis of two types of observations, it ...an be safely assumed that *pancreatic* hormones are ...laborated during embryonic life First, the *effects* of pancreatic insufficiency or ablation *are tempo*...rarily alleviated in the pregnant female Secondly, both insulin and glucagon have been *isolated* from the pancreata of fetal cattle and swine, the *embry*onic organ yielding more hormone per unit weight than adult glands. *See* GLAND. [I.B.K.]

ANATOMY

The pancreas is a more or less developed gland connected with the duodenum. *It can be considered* as an organ characteristic of vertebrates.

Lower vertebrates. In *Amphioxus*, a pancreatic anlage is found in young stages as a thickening of the gut caudal to the liver. According to J. Van Wijhe, it would persist in later stages, but it was not found in the adult. The pancreas of cyclostomes, arising from the gut epithelium or from the liver duct, seems to be purely endocrine; it degenerates in later stages of development.

A true pancreas is found in selachians, with an exocrine portion opening into the intestine and an endocrine portion represented by cellular thickenings of the walls of the ducts

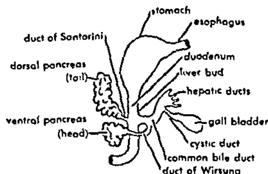


Fig 2. Embryonic development of liver and gallbladder and of dorsal and ventral pancreatic buds before fusion of the latter. (From G C Kent, *Comparative Anatomy of the Vertebrates*, McGraw-Hill, 1954)

Higher vertebrates. Ganoids show a diffuse pancreas—its principal mass lying between the gut and the liver—in which typical islands of Langerhans are observed The pancreas of teleosts is either of the massive or dispersed type Many species, such as the pike, show enormous islands of Langerhans, 10 × 5 mm. from which J McLeod (1922) extracted insulin. The existence of a pancreas in dipneusts, such as *Protopterus*, is doubtful.

The compact pancreas of the amphibians is located in the gastrohepatic omentum and extends towards the hilus of the liver and along the branches of the portal vein It develops from three anlagen, one dorsal and two ventral, the evolution of which varies from one species to another. The *dorsal* anlage would be the only source of endocrine islands The pancreas of reptiles is very similar to that of amphibians, the number of excretory ducts varies from one to three

In birds, the massive pancreas always lies in the duodenal loop. It develops from many dorsal and two ventral thickenings of the duodenal epithelium, one (sometimes two) excretory duct persists The median portion of the dorsal anlage develops into a single mass which subdivides into typical islands

tional condition in the evolution of the vertebrate pancreas. At the junction of the esophagus and intestine, a mass of darkly staining cells forms from *follicles budding off from the intestinal wall.* These follicles lose their connection with the intestinal epithelium and, based on the histological features of their component cells, have been identified with the islet tissue of the higher vertebrates. In the anterior portion of the ammocoete intestine, two bilaterally arranged areas containing cells with a distinctly granular cytoplasm have been described, and it has been suggested that they are homologs of the exocrine portion of the vertebrate pancreas. The granules of these cells exhibit staining properties similar to the zymogen granules of the acinar cells of the pancreas of higher vertebrates. The production of a proteolytic enzyme of the tryptic type, thus similar to one of the digestive enzymes of the

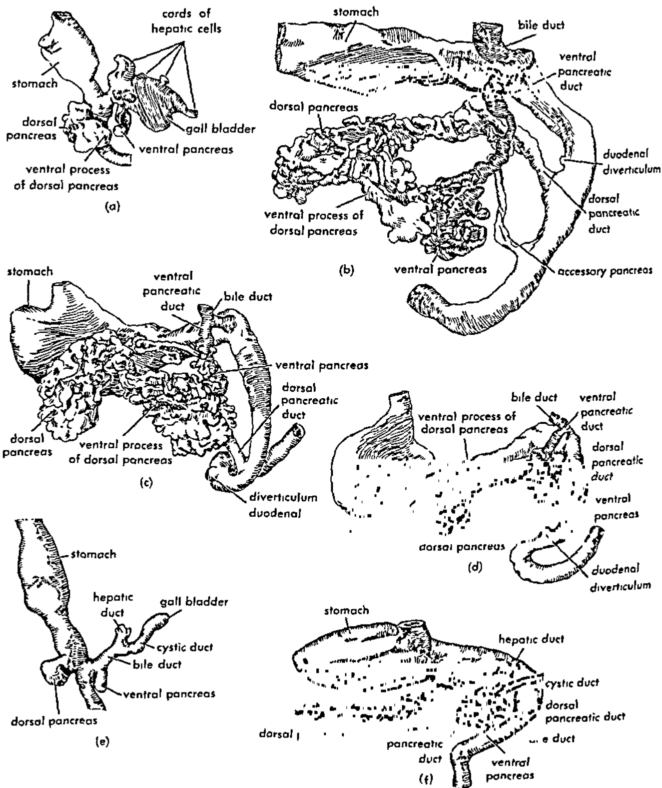


Fig. 1. Reconstruction of the pancreas in various vertebrate embryos. (a) 5.5-mm pig embryo. (b) 20-mm pig embryo. (c) 11-mm rabbit embryo. (d) 10.7-mm cat embryo. (e) 7.5-mm human embryo. (f) 13.6-mm human embryo. (*From F. W. Thyng, Models of the pancreas in embryos of the pig, rabbit, cat and man, Am. J. Anat., 7:488–503, 1907*)

... the anterior portion ... length- ... of an ... estine
rows anteriorly as a pair of blind pouches These pouches are lined with the granular cell type. It is difficult to escape the temptation to speculate that the ammocoete larva represents an evolutionary stage in which the endocrine pancreas has already separated from its original location and that the zymogenlike cells are preparing to follow suit, namely, the condition in the larva of *G. australis.* The next phylogenetic advance is observed in the elasmobranch In these forms, a single dorsal pancreatic primordium and, consequently, the single duct of Santorini develop Thus, phylogenetically as well as ontogenetically, the dorsal pancreas arises first. The first orders in which a pancreas similar to the mammalian organ appears are the ganoids and teleosts

Another observation bespeaking the embryonic origin of the pancreas from the epithelium of the primitive digestive tract is the demonstration that cells with staining properties similar to those of the alpha cells occur in the pyloric portion of the canine stomach Extracts of this region give a positive test for the hormone glucagon

Cell types. Within an islet of Langerhans, at least three distinct granular cell types can be distinguished on the basis of tinctorial differences among the granules demonstrable with polychrome stains Present evidence strongly suggests that the acidophilic alpha cell is concerned with the synthesis of the hormone glucagon and that the basophilic beta cell is the site of *insulin* synthesis. The third cell type, the delta cell, may be a transitional form In addition, a fourth, agranular, cell type may be present in some *species* Although, in most species, each islet contains all of the representative cell types, it has *been* reported that, in the domestic fowl, the individual *islets* may be predominantly alpha-cell or beta-cell islets Considerable variation exists as to the time and order of appearance during development of the *cell* types among the various species In the rat, *the beta* cell can first be discerned in the 18½-day fetus, alpha cells cannot be demonstrated *until* two days postpartum. The alpha cell, on the other *hand*, is the first islet cell type recognized in the chick pancreas, appearing on the eighth day of incubation. On the twelfth day, concomitantly with *the* appearance of large numbers of degenerating cells, *beta cells* can be distinguished As yet, no definitive evidence *is available* which would permit a correlation between the detection of a particular cell type and *the* production of the hormonal product of that cell *type* However, on the basis of two types of observations, it can be safely assumed that *pancreatic* hormones are elaborated during embryonic life *First*, the effects of pancreatic insufficiency or ablation are temporarily alleviated *in* the pregnant female Secondly, both insulin and glucagon have been isolated *from* the pancreas ... cattle and swine, the embryonic organ yielding more hormone per unit weight than adult glands. *See* GLAND. [I R K.]

ANATOMY

The pancreas is a more or less developed gland connected with the duodenum. It can be considered as an organ characteristic of vertebrates.

Lower vertebrates. In *Amphioxus*, a pancreatic anlage is found in young stages as a thickening of the gut caudal to the liver. According to J. Van Wijhe, it would persist in later stages, but it was not found in the adult. The pancreas of cyclostomes, arising from the gut epithelium or from the liver duct, seems to be purely endocrine; it degenerates in later stages of development

A true pancreas is found in selachians, with an exocrine portion opening into the intestine and an endocrine portion represented by cellular thickenings of the walls of the ducts.

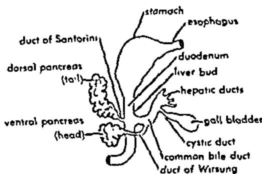


Fig 2 Embryonic development of liver and gallbladder and of dorsal and ventral pancreatic buds before fusion of the latter. (From G C. Kent, *Comparative Anatomy of the Vertebrates, McGraw-Hill, 1954*)

Higher vertebrates. Ganoids show a diffuse pancreas—its principal mass lying between the gut and the liver—in which typical islands of Langer... The pancreas of teleosts is either ...

tracted insulin. The existence of a *pancreas* in dipneusts, such as *Protopterus*, is doubtful.

The compact pancreas of the amphibians is located in the gastrohepatic omentum and extends towards the hilus of the liver and along the branches of the portal vein. It develops from three anlagen, one dorsal and two ventral, the evolution of which *varies* from one species to another. The dorsal anlage would be the only source of endocrine islands The pancreas of reptiles is very similar to that of amphibians; the number of excretory ducts varies from one to three

In birds, the massive pancreas always lies *in the* duodenal loop It develops from many dorsal and *two ventral thickenings* of the duodenal epithelium; one (sometimes two) excretory duct persists. The *median portion* of the dorsal anlage develops into a single mass which subdivides *into typical* islands

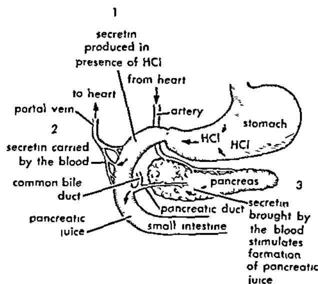


Fig. 3. The path and action of secretin in stimulating production of pancreatic juice. (From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1958)

of Langerhans. A complete ring of pancreatic tissue surrounds the portal vein.

The pancreas of mammals shows the same variations as in the fishes. The extremes are the unique, massive pancreas of man, and the richly branched organ of the rabbit. Usually, the main duct, the duct of Wirsung, opens into the duodenum very close to the hepatic duct. Many rodents have this opening of the pancreatic duct as far as 40 cm from the hepatic duct. In man, the pancreas weighs about 70 g. It can be divided into the head, the body, and the tail. A portion called the uncinate process is more or less completely separated from the head. Accessory pancreases are frequently found anywhere along the small intestine, in the wall of the stomach, and in Meckel's diverticulum. *See* DIGESTIVE SYSTEM

HISTOLOGY

The pancreatic parenchyma is formed by two elements; one is the exocrine tissue of which the secretion empties into the pancreatic ducts and ultimately the duodenum; the other is the endocrine islands whose secretions enter the blood vessels.

Exocrine tissue. The exocrine portion shows tubuloalveolar glands. Each terminal alveolus is called an acinus. The various acini have central cavities, which open into intralobular ducts through narrow intercalated tubes. The interlobular ducts anastomose and ultimately form the main duct of Wirsung, which eventually terminates in the accessory duct of Santorini. The secreting cells of the acini are very similar to those of the salivary glands; they are pyramidal in shape and are exclusively of the serous type. Their basal portion shows numerous filamentous mitochondria. Electron microscope studies provide the basis for understanding this structure. It is rich in ribonucleic acids. The central portion shows the nucleus and, contiguous to it, the zone of Golgi. The apical portion is more or less filled with secretion granules, sometimes reaching the size of 0.9 μ. The cell passes through successive phases of elaboration, excretion, and rest. Functional changes in the mitochondria and the zone of Golgi have been described. The activity of the acini is stimulated by secretin as well as by pilocarpine.

Endocrine tissue. The endocrine portion shows cellular masses called islands or islets of Langerhans, in which the cellular cords or masses are more or less isolated by irregular spaces filled with connective tissue and blood capillaries. With suitable staining techniques, one can identify various types of cells in these islands. The two main elements are the alpha and the beta cells, which exist usually in the proportion of 1:4. Other types of cell are found occasionally, such as the gamma cells in the pancreas of the guinea pig, the delta cells mainly found in man, the epsilon cells described in the opossum, and the X cells which have been found in the horse and which may be identical to the elements described in the enormous islands of certain birds. Among other differential characters of the various cell types may be mentioned the existence of alkaline phosphatase in the alpha cells and the presence of zinc in the beta elements. The number and the distribution of the various types of cell, as well as their cytological features, vary considerably with experimental conditions.

Between the grapelike exocrine portion with its ducts and the islands of Langerhans, one observes connective tissue septa, numerous blood vessels, and nerves. Nervous ganglia are variable in number; they are sometimes closely associated with islands, forming the sympatheticoinsular associations. Tactile corpuscles of Pacini are frequent in the cat's pancreas.

PHYSIOLOGY

The pancreatic juice which is carried to the duodenum is a slightly alkaline liquid containing trypsinogen, which, when activated, causes the hydrolysis of the proteins into amino acids, amylase and maltase, which act on the glucides, and lipase, which causes the hydrolysis of fatty substances. Nothing definite is known about the cellular origin of these various digestive ferments. The intense stimulation of the pancreatic secretion after ingestion of food is considered to be the result of a nervous reflex originating in the mouth and transmitted to the pancreas by way of the pneumogastric nerves. Pancreatic secretion is also stimulated by direct introduction of acids and fats into the duodenum, causing the liberation of a hormone called secretin into the blood stream to stimulate the exocrine secretion.

Endocrine function. In 1889, J. von Mering and O. Minkowsky were the first to remove completely the pancreas of a dog, thereby causing true diabetes. This experiment is easily performed in most animals with the same results. In the rabbit, however, the complete removal of the gland is difficult, because of its ramified constitution. In birds, pancreatectomy is followed only by a very transient

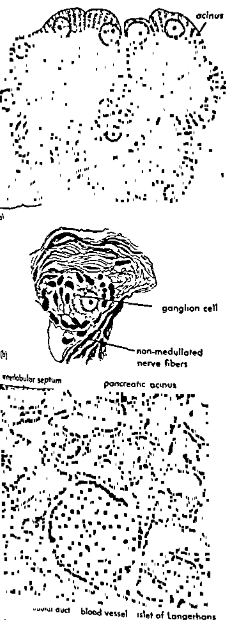


Fig. 4. (a) Centroacinar cells in the pancreas of a guinea pig (from J F Nonidez and W F. Windle, *Textbook of Histology*, 2d ed, McGraw-Hill, 1953) (b) A nerve ganglion cell in the human pancreas (from J L. Bremer and H L Weatherford, *Textbook of Histology*, 6th ed, Blakiston, 1944). (c) A section through the pancreas of the rat The islet of Langerhans is a gland of internal secretion, whereas the surrounding acinar tissue forms an exocrine gland (from C. D Turner, *General Endocrinology*, 2d ed, Saunders, 1955).

hyperglycemia F. Banting and C Best (1922) prepared pancreatic extracts which were able to prevent the lethal effects of pancreatectomy The same effect was obtained with extracts from pancreas in which after ligature of the duct of Wirsung the exocrine portion of the gland had disappeared. The insular origin of the active factor, called ins... De Meyer in 1909, was proved A considerable amount of material, histological, embryological, experimental, and pathological, enables one to differentiate the role of the two main cellular constituents of the islands of Langerhans, namely, the alpha and the beta cells *See* GLAND; ENDOCRINE GLAND.

Effect of alloxane. The beta cells secrete insulin, that is, the hypoglycemic factor. The injection of alloxane, according to J. Dunn and N. McLetchie, 1943, causes a selective destruction of the beta cells and the appearance of the various symptoms of diabetes. The mechanism of the diabetogenic action of alloxane is still a matter of discussion. Repeated injections of the extract of the anterior lobe of the pituitary gland cause alterations of the beta cells and diabetes, the severity of this diabetes is proportional to the degree of degranulation of the cells. The histological study of the pancreas of men who had suffered from diabetes frequently shows marked alterations of the beta elements Certain types of pancreatic adenomas found in the human are essentially constituted by beta cells, and they are associated with hyperinsulinism and hypoglycemia As further arguments in favor of the beta-cell origin of insulin, the damaging effect of dithizone on the dog's beta cells, associated with diabetes and with the disappearance of zinc from the cells, may be mentioned Dialuric acid has identical effects, followed by regeneration of beta cells from acinous tissue. Various substances studied by Kadota and Midokawa cause alterations of which the specificity is not evident.

HGF factor. Intravenous injections of insulin cause a transitory hyperglycemia, soon followed by the typical hypoglycemia This was attributed to the existence in the injected insulin or pancreatic extract of another substance, called the hyperglycemic factor or glucagon This factor has been purified, and its principal effect is to cause hyperglycemia through a process of glycogenolysis. Thus, it is known as the hyperglycemic-glycogenolytic factor or HGF Besides the upper two-thirds of the gastric mucosa of dogs and rabbits, the pancreas is the only organ yielding HGF The insular origin of this factor is shown by the strong correlation between glucagon content and density of insular tissue, the highest content being found in fetal pancreas and the tail part of the adult pancreas. Atrophy of the acinous tissue after ligation of the ducts and degeneration of the beta cells following injection of alloxane do not affect the glucagon content

Experimental aspects. The alpha-cell origin of the HGF is strongly supported by experiments It has been shown that intoxication by the mushroom *Amanita phalloides* causes destruction of the alpha cells, associated with hepatic steatosis and hypoglycemia It has been demonstrated also that repeated injections of cobaltous chloride cause a complete degranulation of the alpha cells, which lasts only a few days Identical effects have been obtained with other cobaltous salts under various experimental conditions and with nickel salts During the phase of degranulation of the alpha cells,

the glucagon content of the pancreas is decreased by about two-thirds. This correlation between the histological observations and the biochemical determinations is a positive argument in favor of the role of the alpha cells in the secretion of HGF.

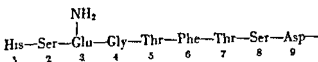
Increase in the number and in the size of the islands has been observed in many experimental conditions, such as fasting, pregnancy, biliary derivation, partial pancreatectomy, castration, and stimulation of the autonomic nervous system.

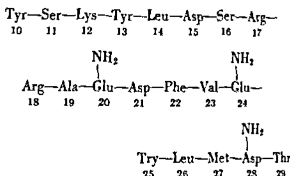
In addition to insulin and glucagon, the pancreas is the source of the lipocaic factor, which prevents the hepatic alterations observed after pancreatectomy. Its origin is unknown. Vagotonine, which regulates the autonomic tonus, might be of insular origin also. *See* METABOLIC DISORDERS; METABOLISM. [E.V.C.]

BIOCHEMISTRY

Hormones. Scattered through the pancreas are isolated masses of endocrine structures, known as the islets of Langerhans. Two types of cells appear in the islets, alpha cells and beta cells These are the sources of two hormones, insulin from the beta, and glucagon, also known as the hyperglycemic factor (HGF), from the alpha cells The former is a hormone which influences carbohydrate metabolism, enabling the organism to utilize sugar (*see* INSULIN) The latter accelerates the conversion of liver glycogen, the form in which carbohydrate is stored in liver and muscles until needed by the body, into glucose, the principal sugar used by the body to meet its energy requirements. Thus, glucagon elevates the blood sugar level, and its effects are the opposite of insulin, so that the two hormones together maintain the sugar metabolism of the body in balance. When the level of sugar in the blood becomes too low, the secretion of glucagon is stimulated

Glucagon. The existence of glucagon in pancreatic extracts was first demonstrated by O Kimball and J. Murlin in 1923. It is frequently present in commercial preparations of insulin as an impurity; the administration of such insulin preparations causes an initial rise in blood sugar, followed by the lowering of blood sugar (hypoglycemia) characteristic of insulin action. In 1955, A. Staub, O. Behren, and their coworkers obtained glucagon in crystalline form; two years later, the same group of investigators established the complete structural formula of the hormone as a single polypeptide chain of 29 amino acids with the following sequence, where His is histidine; Ser, serine; Glu, glutamic acid; Gly, glycine; Phe, phenylalanine; Thr, threonine; Asp, aspartic acid; Tyr, tyrosine; Lys, lysine; Leu, leucine; Arg, arginine; Ala, alanine; Val, valine; Try, tryptophan; Met, methionine:





Pancreatectomy. Within 24 hours after the surgical removal of the pancreas, hyperglycemia, or elevation of blood sugar, is detectable, followed by glycosuria, or excess sugar secreted in the urine. Without the pancreas, the organism is unable to store or to oxidize sugar. This disturbance of sugar metabolism is also accompanied, when there is pancreas deficiency, by defective fat metabolism The organism, in an attempt to compensate for the loss of its primary source of energy, sugar, increases the catabolism or breakdown of fats. When the stores of carbohydrate in the body are depleted, ketone bodies, acid products which are responsible for the metabolism of fats, increase in both the blood and the urine, with serious consequences. Accumulation of these acid products leads to a depletion from the blood of the alkali reserve that balances them in the normal organism. As a consequence, the organism is unable to remove accumulating carbon dioxide in the blood, resulting in acidosis, or air-hunger, which leads to coma and death. These symptoms are known to be associated with diabetes mellitus, which, in man, is characterized by a marked loss of weight, hyperglycemia, glycosuria, ketonemia, ketonuria, and polyuria.

Another pancreatic disorder, which has been observed in rare instances of tumors of the islet tissue, is known clinically as hyperinsulinism. The disease, associated with low blood sugar and indistinguishable from the consequences of insulin overdosage, is caused by excessive production of insulin by the tumor. [C.H.L.]

Pancreas disorders

The pancreas is composed of two different tissues which have dissimilar structure and function, and hence are susceptible to different diseases. The bulk of the gland is composed of the acinar tissue, secreting digestive enzymes which are drained into the first portion of the small intestine through the pancreatic duct. These enzymes, lipase, amylase, and trypsin, facilitate the digestion of fats, carbohydrates, and proteins by the intestine. The other portion of the gland consists of small nests of endocrine cells, the so-called islets of Langerhans. They secrete the hormone insulin which passes directly into the blood stream and regulates the metabolism of glucose by the tissues throughout the body. For a discussion of the normal structure and function of this organ, *see* PANCREAS

Cystic fibrosis. Cystic fibrosis of the pancreas is a hereditary metabolic disorder of unknown cause

affecting various glandular tissues throughout the body, including the acinar portion of the pancreas. In the pancreas it is characterized by obstruction of the ducts by thick, viscid, mucoid secretions. The ducts and glands become dilated behind the obstruction, eventually atrophy and replacement by connective tissue occur. Deficiency of pancreatic enzymes in the intestine results in impaired digestion of food-stuffs, fatty diarrhea, malnutrition, and vitamin deficiency. Similar obstruction may be found concomitantly in the bile ducts of the liver and in the lungs, where thick, mucoid secretions block the small branches of the bronchial tree. Careful medical management successfully compensates for the pancreatic deficiency, but the changes in the lungs result in recurrent episodes of pneumonia, in bronchiectasis, and in death in early childhood. See BRONCHIAL DISORDERS; PNEUMONIA.

Pancreatitis. Acute pancreatitis, or inflammation of the gland, is a serious illness in which there occurs necrosis (death) of varying amounts of the pancreas and the surrounding tissues. It is initiated by the escape of the digestive enzymes from the ducts or glands. Once outside the glands or ducts, these enzymes become activated, digest the pancreatic tissue producing them, and allow further escape of enzymes. Dissolution of the walls of blood vessels by trypsin may produce extensive hemorrhage. The clinical effects are severe pain and a state of deep shock which may be fatal. The primary cause of acute pancreatic necrosis is a subject of great controversy. One theory is that obstruction of the ampulla of Vater by gallstones or muscle spasm causes bile to be regurgitated from the common bile duct into the pancreatic duct, where it has adverse effects. More widely accepted is the view that in response to the stimulus of alcohol or a large meal, pancreatic secretion exceeds the drainage capacity of the ducts. Pressure within the ducts becomes sufficiently high to cause rupture. Whatever the cause, the disease is unduly common in individuals who overindulge in alcohol. Diagnosis is facilitated by identification of an excessive quantity of amylase in the blood or urine. Survival may be associated with large deposits of calcium in the sites around the pancreas where fatty tissues were destroyed. Although a recognized instance of acute pancreatitis represents a severe and hazardous illness, many minor unrecognized subclinical cases also occur. These may masquerade as minor digestive upsets, yet produce appreciable scarring of the gland. See DEATH; GALLSTONES.

Stones or concretions may develop in pancreatic ducts, usually from unknown causes. By obstructing the ducts, they may cause a wide dilatation of ducts and glands behind the obstruction with eventual atrophy and fibrosis. [illegible] pancreatitis or [illegible] insufficiency rarely results from such obstruction, and islet tissue usually persists even in the face of extensive fibrous replacement of the acinar tissue. Cysts of the pancreas may [illegible] tal malformations [illegible] are of little [illegible]

that they press on adjacent organs or simulate tumors and necessitate surgical exploration.

Diabetes. Diabetes mellitus is the major disease associated with the insulin secreting islet cells of the pancreas. The firm establishment of the role of insulin and the pancreas in this disease was one of the great milestones in medical progress, and for it F. Banting and C. Best received a Nobel Prize. A relative or absolute deficiency of insulin, with resultant diabetes mellitus, is reflected physiologically by an inability of the tissues to use glucose in their metabolism. This sugar consequently accumulates in excess in the blood and is excreted in the urine. The tissues rely on excessive metabolism of proteins and fats as substitutes for sugar. This results in depletion of tissue proteins and in the accumulation of acid waste products from incomplete metabolism of fats. These waste products, called ketone bodies, are responsible for episodes of diabetic coma, a common cause of death of these individuals before insulin therapy. Diabetic patients are also unduly susceptible to infections, severe arteriosclerosis, gallstones, and a unique and serious kidney disorder called intercapillary glomerulosclerosis. Fortunately the administration of insulin greatly minimizes all of these possible complications, and diabetics now can be expected to live long and active lives.

Cause. The cause of diabetes is still not well established beyond the fact that insulin secretion is relatively or absolutely insufficient. In rare cases only is there sufficient destruction of islet tissue to account for the disease. In many cases there are minor anatomical changes in the islands of Langerhans, insufficient in extent to explain the insulin deficit. And in still other diabetics (as many as 25%) no anatomical abnormalities can be found in the islet cells. It is postulated that in the usual case of diabetes mellitus the fault is a disturbance of the normal interrelationship between the secretion of these islet cells of the pancreas and the secretions of the pituitary or the adrenal glands.

Tumors. Tumors of the acinar tissue of the pancreas which are of any significance are malignant in nature. The manifestations of adenocarcinoma of the pancreas depend to a considerable extent on the location of the tumor in the gland. Tumors arising in the head, or that portion of the gland adjacent to the duodenum, cause early obstruction of the common bile duct and consequent jaundice. In more distal portions of the gland, back pain is the major symptom and is often late in appearing. In either situation, surgical removal of the tumor is difficult and rarely successful; the patient eventually dies from spread of the tumor into the liver and other vital organs.

Benign adenomas and less commonly adenocarcinomas may arise from islet tissue. Either tumor may produce insulin in excessive quantities. This is manifested clinically by episodes of fainting due to unduly low concentrations of sugar in the blood. The effects on the metabolism of the cells of the brain may be fatal. If the tumor is

the glucagon content of the pancreas is decreased by about two-thirds. This correlation between the histological observations and the biochemical determinations is a positive argument in favor of the role of the alpha cells in the secretion of HGF.

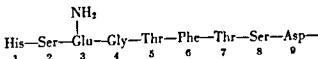
Increase in the number and in the size of the islands has been observed in many experimental conditions, such as fasting, pregnancy, biliary derivation, partial pancreatectomy, castration, and stimulation of the autonomic nervous system.

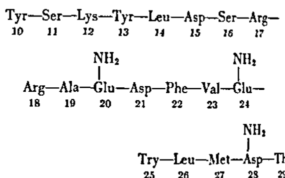
In addition to insulin and glucagon, the pancreas is the source of the lipocaic factor, which prevents the hepatic alterations observed after pancreatectomy. Its origin is unknown. Vagotonine, which regulates the autonomic tonus, might be of insular origin also. *See* METABOLIC DISORDERS; METABOLISM. [E.V.C.]

BIOCHEMISTRY

Hormones. Scattered through the pancreas are isolated masses of endocrine structures, known as the islets of Langerhans. Two types of cells appear in the islets, alpha cells and beta cells. These are the sources of two hormones, insulin from the beta, and glucagon, also known as the hyperglycemic factor (HGF), from the alpha cells. The former is a hormone which influences carbohydrate metabolism, enabling the organism to utilize sugar (*see* INSULIN) The latter accelerates the conversion of liver glycogen, the form in which carbohydrate is stored in liver and muscles until needed by the body, into glucose, the principal sugar used by the body to meet its energy requirements Thus, glucagon elevates the blood sugar level, and its effects are the opposite of insulin, so that the two hormones together maintain the sugar metabolism of the body in balance. When the level of sugar in the blood becomes too low, the secretion of glucagon is stimulated

Glucagon. The existence of glucagon in pancreatic extracts was first demonstrated by O. Kimball and J. Murlin in 1923. It is frequently present in commercial preparations of insulin as an impurity, the administration of such insulin preparations causes an initial rise in blood sugar, followed by the lowering of blood sugar (hypoglycemia) characteristic of insulin action In 1955, A. Staub, O. Behren, and their coworkers obtained glucagon in crystalline form; two years later, the same group of investigators established the complete structural formula of the hormone as a single polypeptide chain of 29 amino acids with the following sequence, where His is histidine; Ser, serine; Glu, glutamic acid; Gly, glycine; Phe, phenylalanine; Thr, threonine; Asp, aspartic acid; Tyr, tyrosine; Lys, lysine; Leu, leucine; Arg, arginine; Ala, alanine; Val, valine; Try, tryptophan; Met, methionine:





Pancreatectomy. Within 24 hours after the surgical removal of the pancreas, hyperglycemia, or elevation of blood sugar, is detectable, followed by glycosuria, or excess sugar secreted in the urine. Without the pancreas, the organism is unable to store or to oxidize sugar. This disturbance of sugar metabolism is also accompanied, when there is pancreas deficiency, by defective fat metabolism. The organism, in an attempt to compensate for the loss of its primary source of energy, sugar, increases the catabolism or breakdown of fats. When the stores of carbohydrate in the body are depleted, ketone bodies, acid products which are responsible for the metabolism of fats, increase in both the blood and the urine, with serious consequences. Accumulation of these acid products leads to a depletion from the blood of the alkali reserve that balances them in the normal organism. As a consequence, the organism is unable to remove accumulating carbon dioxide in the blood, resulting in acidosis, or air-hunger, which leads to coma and death. These symptoms are known to be associated with diabetes mellitus, which, in man, is characterized by a marked loss of weight, hyperglycemia, glycosuria, ketonemia, ketonuria, and polyuria

Another pancreatic disorder, which has been observed in rare instances of tumors of the islet tissue, is known clinically as hyperinsulinism. The disease, associated with low blood sugar and indistinguishable from the consequences of insulin overdosage, is caused by excessive production of insulin by the tumor [C.H.L.]

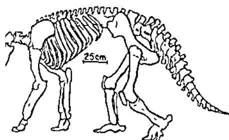
Pancreas disorders

The pancreas is composed of two different tissues which have dissimilar structure and function, and hence are susceptible to different diseases. The bulk of the gland is composed of the acinar tissue, secreting digestive enzymes which are drained into the first portion of the small intestine through the pancreatic duct These enzymes, lipase, amylase, and trypsin, facilitate the digestion of fats, carbohydrates, and proteins by the intestine The other portion of the gland consists of small nests of endocrine cells, the so-called islets of Langerhans They secrete the hormone insulin which passes directly into the blood stream and regulates the metabolism of glucose by the tissues throughout the body. For a discussion of the normal structure and function of this organ, *see* PANCREAS.

Cystic fibrosis. Cystic fibrosis of the pancreas is a hereditary metabolic disorder of unknown

five toes each. The pantodont dentition is iced, with large canines, obliquely crested and V-shaped premolars The skull is usually low and wide, with a strong crest above the case Horns are absent.

Pantodonts are divided into four families: Archaeolambdidae (*Archaeolambda*, early Eo- Mongolia), (2) Coryphodontidae (*Pantolambda*, *Titanoides*, *Caenolambda*, *Coryphodon*, *Hypoceras*, *Procoryphodon*, *Hypercoryphodon*, late Paleocene-early Eocene of North America, Eocene of Europe, late Paleocene-middle Eocene of Mongolia); (3) Barylambdidae (*Barylambda*, *Haplolambda*, late Paleocene, North America); and (4) the problematical Pantolambdodontidae (*Pantolambdodon*, late Eocene, Mongolia)



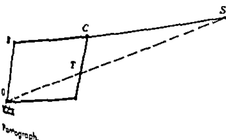
Skeleton of *Barylambda*, a late Paleocene pantodont (after B Patterson)

Pantolambda, from the middle Paleocene of North America, is the most primitive pantodont. The ancestry of the order is still unknown but may be near the genus *Deltatherium* in the creodont carnivorans. See CARNIVORA FOSSILS

Coryphodon, the first-discovered pantodont (1840), was dredged from submarine rock outcrops in the English Channel This best known and most commonly found genus is recorded in the late Paleocene and early Eocene of North America and early Eocene of Europe Pantodonts became extinct without giving rise to any known modern or late Cenozoic descendants [M C.MC.]

Pantograph

A four bar parallel linkage, with no links fixed, used as a copying device for generating geometrically similar figures, larger or smaller in size, within the limits of the mechanism The figure traced by point T, as shown in the illustration, will be similar to that generated by point S This similarity results because points T and S will always lie on the straight line $\overline{OTS}$; triangles OBS and TCS are always similar, because lengths $\overline{OB}$, $\overline{BS}$, $\overline{CT}$, and $\overline{CS}$ are constant, and $\overline{OB}$ is always parallel to $\overline{CT}$. Distance $\overline{OT}$ always maintains a constant proportion to distance $\overline{OS}$, because of the similarity of the above triangles. Numerous modifications of the pantograph as a copying device have been made.



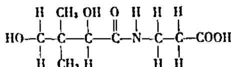
Pantograph.

James Watt applied the pantograph as a reducing motion in his beam engine (see STRAIGHT LINE MECHANISM). The pantograph may also serve as a reducing motion for an engine indicator; S is attached to the engine crosshead while the indicator cord is attached to T.

The collapsible parallel linkage used on electric locomotives and rail cars to keep a collector bar or wheel in contact with a trolley wire is also called a pantograph. [E.S F]

Pantothenic acid

A member of the B vitamin group with the structural formula



$$\mathrm{HO{-}\underset{H}{\overset{H}{C}}{-}\underset{CH_3}{\overset{CH_3}{C}}{-}\underset{H}{\overset{OH}{C}}{-}\overset{O}{\overset{\|}{C}}{-}\overset{H}{N}{-}\underset{H}{\overset{H}{C}}{-}\underset{H}{\overset{H}{C}}{-}COOH}$$

It is a light yellow, viscous oil which is readily soluble in H_2O It is usually obtained as the calcium salt It is stable in neutral solution but is decomposed by hot acid or alkali to pantoic acid or its lactone and β-alanine Chemical methods for estimating the amount of pantothenic acid are unsatisfactory Rat and chick growth assays have been used successfully, but microbiological assays using bacteria or yeast are more widely used

Figures for the pantothenic acid content of foods are not as reliable as those for some other vitamins Pantothenic acid is widely distributed, and liver, kidneys, fresh green vegetables, and egg yolks are among its best sources Losses of the vitamin during cooking are minimal, as it is present in stable conjugated form in food

No definite pathologic lesions due to a specific pantothenic acid deficiency have been reported in man In prisoners of war and malnourished people of Asia, a "burning feet" syndrome has been described which responds well to pantothenic acid. All animal species studied require a dietary source of pantothenic acid, and it is probable that man does too The deficiency state in animals has resulted in poor growth, dermatitis, adrenal necrosis and hemorrhage, kidney damage, blood dyscrasias, and myelin degeneration of brain and spinal cord In some species pantothenic acid deficiency has caused graying of the hair, which has resulted in unfounded claims concerning the use of pantothenic acid preparations to prevent graying in humans

Pantothenic acid exists in enzyme systems as coenzyme A (CoA) CoA is a nucleotide consisting

benign, simple surgical removal is curative. See NEOPLASIA. [M.R.H.]

Bibliography: S. L. Robbins, *Textbook of Pathology*, 1957; H. A. Smith and T. C Jones, *Veterinary Pathology*, 1957.

Panda

Any of two genera of large Asiatic carnivores of the family Procyonidae, related to the raccoon. Pandas look more like toy teddy bears than real animals; they are among the most popular of all zoo animals. The Himalayan panda, *Ailurus* sp., is



The giant panda, *Ailuropoda melanoleucus*; length to 4 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

not as well known as the giant panda, *Ailuropoda melanoleucus*, which is native to western China. near the border of Tibet. The latter is about 5 ft tall and is white with black legs, black eye spots, and black ears. See CARNIVORA [J.D.B.]

Pandanales

An order of the plant subclass Monocotyledoneae consisting of three families: Typhaceae (cattail family) with one genus, Sparganiaceae (burreed family) with one genus, and Pandanaceae (screwpine family) with three genera. Cattails (12 species) are found in marshes throughout the world Burreeds occur in marshes and along shores throughout the Northern Hemisphere and New Zealand The screwpines, so called because of their conspicuously spiraled leaves, are found from Africa through Indonesia to Australia The flowers in this order are naked, that is, they have no perianth (floral envelope). See EMBRYOPHYTA; MONOCOTYLEDONEAE; PLANT KINGDOM [P.D.S.]

Panel heating and cooling

A system in which the heat-emitting and -absorbing means is the surface of the ceiling, floor, or wall panels of the space which is to be environmentally conditioned.

The heating or cooling medium may be air, water, or other fluid circulated in air spaces, conduits, or pipes within or attached to the panel structure. For heating only, electric current ma[y] [f...] through resistors in or on the panels.

The most commonly used medium is warm [...] cold water, circulated within steel pipes embed[ded] in concrete floors or ceilings or in plaster ceil[ings].

Heat transfer. Heat energy is transmitted f[rom] a warmer to a cooler mass by conduction, con[vec]tion, and radiation (see HEAT TRANSFER). R[adi]ant heat rays are emitted from all bodies at t[em]peratures above absolute zero. These rays p[ass] through air without appreciably warming it, [but] are absorbed by liquid or solid masses and [in]crease their sensible temperature and heat conte[nt].

The heat output from heating means compri[ses] both radiation and convection components in va[ry]ing proportions In panel heating systems, es[pe]cially the ceiling type, the radiation compon[ent] predominates. Heat interchange follows the Stef[an]-Boltzmann law of radiation; that is, heat trans[fer] by radiation between surfaces visible to ea[ch] other varies with the fourth power of the absol[ute] temperature difference between the surfaces, a[nd] is transferred from the surface with the higher te[m]perature to the surface with the lower temperatu[re].

The skin surface temperature of the human bo[dy] under normal conditions varies from 87 to 95 [...] and is modified by clothing and rate of metaboli[sm]. The presence of radiating surfaces above these te[m]peratures heats the body whereas those below p[ro]duce a cooling effect. See RADIANT HEATING.

Use for cooling. When a panel system is us[ed] for cooling, the dew-point temperature of the a[m]bient air must remain below the surface tempe[ra]ture of the heat-absorbing panels to avoid conde[n]sation of moisture on the panels. In regions whe[re] the maximum dew-point temperature does not e[x]ceed 60°F, or possibly 65°F, as in the Pac[ific] Northwest and the semi-arid areas between the Ca[s]cade and Rocky Mountains, ordinary city wat[er] provides radiant comfort cooling

Where higher dew points prevail, it is necessa[ry] to dehumidify the ambient air. Panel cooling e[f]fectively prevents the disagreeable feeling of co[ld] air blown against the body and minimizes the o[c]currence of summer colds.

Records of fuel consumption show that pan[el] heating systems save 30–50% of the fuel costs [of] ordinary heating systems, because lower ambien[t] air temperatures produce comfort; air temper[a]tures within the room are practically uniform an[d] not considerably higher at the ceiling as in radia[tor]- and especially in convector-heated interior[s]. See COMFORT CONTROL [F.L.W.]

Bibliography· American Society of Heating an[d] Ventilating Engineers, *Heating, Ventilating, Ai[r] Conditioning Guide*, vol 37, 1959; F E. Giesecke, *Hot-water Heating and Radiant Heating and Radi[ant] Cooling*, 1947

Pantodonta

A group of extinct, large, hoofed herbivorous quad[ru]pedal mammals of the early Cenozoic of Europe, North America and northeastern Asia. The feet o[f] these poorly known forms are semigraviportal and

er of food plants including cabbage, broccoli, els sprouts, cauliflower, collards, kohlrabi, tur-rutabaga, kale, horseradish, radish, mustard, ress

e poppy family (Papaveraceae) with regular, g flowers usually having numerous stamens, is econd largest. The fruit is a capsule. Several es are ornamentals and *Papaver somniferum* is urce of opium and morphine.

mbers of the remaining four families have ir-lar flowers. They include a few well-known mentals such as bleeding heart, spider flower, mignonette Ben oil, a nondrying oil used in icating watches, is obtained from the seed of the le fruits of the horseradish tree, *Moringa olei-*

d in this article; see POPPY; see also DICOT-YONEAE; EMBRYOPHYTA; PLANT KINGDOM; ETABLE GROWING. [P D S.]

per and paper products

term *paper* is traditionally applied to felted matted sheets of cellulose fibers, formed on a wire screen from a dilute water suspension, bonded together as the water is removed and sheet is dried In present-day usage, paper may o include sheet materials produced from other es of fibers (particularly mineral or synthetic) med and bonded by other means

Paper was employed originally as a medium for ting As such, it has played a significant role in e development of civilization and culture. Writ-g papers today, however, account for a relatively all proportion of the output of the paper in-stry Of the paper production of nearly 31,000,-0 tons in 1958 in the United States, about 1,600,-0 tons was for writing and other fine papers, 00,000 newsprint, 2,000,000 tissue, 4,000,000

...cellulose fibers have oc-cupied the dominant position for nearly 2000 years. Other fibrous materials, particularly some of the synthetic high polymers, are assuming increasing importance in papermaking, but it is predicted that cellulose will continue to play the key role for many years Numerous other materials are used in processing and converting operations, but they will be mentioned only incidentally in this article

Wood pulp Wood is the primary source of cellu-lose fibers for papermaking Before use, the wood must be reduced to the fibrous state; this opera-tion is called pulping At present, commercial pulping operations are of three principal types: mechanical, full chemical, and semichemical See WOOD FIBER PRODUCTS

Mechanical pulping, as the name implies, in-volves the reduction of wood to the fibrous state by purely mechanical means. Logs, of either 2- or 4-ft length, are ground into pulp by means of large revolving grindstones. The logs are held trans-versely against the surface of the stone, with the axis of the log parallel to the stone axis. Water is sprayed against the stone to control the tempera-ture and to carry away the resulting pulp.

Except for a few water-soluble components, all the constituents of the wood are present in the groundwood pulp. Thus, the yield of pulp from wood may be nearly 95%. Principally, coniferous wood species have been used for groundwood, al-though grinding of deciduous species is practiced to some extent.

Full chemical pulping employs chemical re-agents to effect a separation of the cellulose fibers from the other wood components. Wood chips are cooked with suitable chemicals in aqueous solution, usually at elevated temperatures and pressures. The object is to dissolve the lignin and other extraneous compounds, leaving the cellulose intact and in fibrous form. Although there is always some cellu-lose degradation, the objective can be realized to a commercially satisfactory degree through the use of a variety of chemical reagents. Pulp yields are usually about 50% of the wood weight.

The sulfite process has been used largely for pulping spruce, fir, and hemlock, although the pines and hardwoods are also being reduced by this proc-ess under certain conditions. The cooking acid con-sists of a solution of an alkali or alkaline-earth bisulfite, also containing free sulfurous acid. So-called calcium-base cooking acids have been most widely used in the past, but there is today a signifi-cant shift toward the use of magnesium, ammonia, or sodium as the base. These soluble bases are advantageous in certain instances if liquor-re-covery operations are to be practiced. Sulfite pulps are relatively light in color, are easily bleached, have moderately good strength properties, and are widely used in fine papers

The kraft or sulfate process is the most exten-sively employed Here the active pulping ingredi-ents are sodium hydroxide and sodium sulfide, in an obviously strongly alkaline solution Almost any wood species can be pulped by this process. Stand-ard in kraft pulping is a liquor-recovery cycle, in which the dissolved organic constituents in the spent pulping liquors are burned for steam genera-tion, and the inorganic pulping chemicals are re-covered and reused Kraft pulps are dark in color, difficult to bleach, and very strong In the un-bleached state, they are employed widely in the container field, and they may be bleached and used in fine papers.

The full chemical pulping technique, the soda process, is now largely of historical interest.

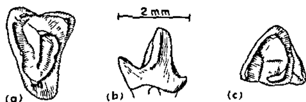
Semichemical pulping, although relatively new, is of great current interest Several processes are being employed, in which mild chemical action, followed by mechanical attrition, is used Among these are the neutral sulfite, semikraft, and cold soda processes A variant is the chemi-groundwood process, in which the logs are thoroughly impreg-nated with a hot chemical solution before grinding These processes are employed largely, but not ex-clusively, on deciduous wood species. Yields may

acid, and β-mercaptoethylamine. The enzyme can be acetylated at its sulfhydryl-containing end, and this acetyl CoA is the active acetate reported in a variety of biochemical processes. CoA is of prime importance in the metabolism of acetyl groups and is therefore active in the metabolism of fats, carbohydrates, and many other substances. It is important in the synthesis of steroid hormones, cholesterol, acetylcholine, fatty acids, and some amino acids. CoA is not only an acetyl carrier; it transports longer acyl chains, and as succinyl CoA is involved in the synthesis of the pyrrole rings of hematin

The factors affecting the requirement for pantothenic acid are probably similar to those altering the needs for the other B vitamins. Pantothenic acid has been reported to improve the reactions of young men to stress This fact and the effect of the deficiency on the adrenals suggest the possibility that stress increases the pantothenic acid requirement The widespread occurrence of pantothenic acid assures protection against the deficiency state under most conditions Most Americans eat 10 mg of pantothenic acid per 2500 cal of good diet. The daily requirement is probably about 3–5 mg. *See* BIOASSAY, COENZYME; ENZYME; VITAMIN. [S.N.G.]

Pantotheria

The pantotheres were one of the most abundant types of carnivorous or insectivorous Jurassic mammals. They have been found in the Middle and Upper Jurassic of England and the Upper Jurassic of North America and Africa. The lower molars of the pantotheres consisted of three main cusps arranged in a triangle (the trigonid) with a small heel or ledge (the talonid) at the base of the posterior side of the trigonid The cusps of the pantothere trigonid are probably homologous with the paraconid, protoconid, and metaconid of later mammals. The upper molars are unique, consisting of two large main cusps, one near or external to the center of the crown and the other at the internal corner, and a variable number of accessory cusps. The primitive cheek tooth formula was probably four premolars and eight molars. The proportions of the jaw were variable, but a distinct angular process was always present.



Pantotheres. (a) Occlusal view of an upper molar of *Melanodon*. (b, c) Internal and occlusal views of a lower molar of *Dryolestes*. (After G. G. Simpson, 1929)

The seven molars of the Middle Jurassic genus *Amphitherium*, the oldest pantothere and only member of the Amphitheriidae, have relatively long talonids. The long talonid and the large number of molars indicate that *Amphitherium* was probably not far removed from the primitive pantothere condition.

The Upper Jurassic pantotheres are easily separated into two families which show diametrically opposed modes of specialization. In the Paurodontidae the molar talonid is not shortened, the number of molars is reduced to a maximum of four, and the jaw tends to be short and stout. In the Dryolestidae the molar talonid is shortened, the number of molars is always seven or more, and the lower jaw tends to be long and slender.

Many workers believe that the order Symmetrodonta was ancestral to the Pantotheria, but this has been vigorously disputed. For some time the pantotheres have been considered to be ancestral to the later mammals, the so-called placentals and marsupials. The morphology of the pantothere lower molar favors this relationship, but until recently the unique morphology of the upper molars proved to be a stumbling block in relating the groups.

Bryan Patterson recently described a collection of mammal teeth found in the Middle Cretaceous Trinity sands of Texas. In addition to a symmetrodont and a triconodont he described a type of mammal which appears to be intermediate between the pantotheres and the modern types of mammals. This study led Patterson to believe that the large internal cusp of the pantothere upper molar was homologous with the paracone of later mammals and that the other main cusp of the pantothere molar was reduced to a weak external stylar cusp or lost in later mammals. According to this theory the major changes involved in the transition from the pantothere molar to the molar of later mammals were (1) the addition of the protocone and consequent modifications of the upper molar and (2) small but important modifications of the talonid of the lower molars.

Endotherium, known only from a single fragmentary lower jaw found in beds of Late Jurassic or Early Cretaceous age in Manchuria, has uncertain relationships, but it may be related to the Forestburg mammals.

Patterson's conclusions may be the explanation for the long-standing belief that the pantotheres were the ancestors of the later orders of mammals It is fairly certain, however, that *Endotherium* and the Forestburg mammals represent a group transitional between pantotheres and later mammals. *See* THERIA. [W.A.CL.]

Papaverales

An order of the plant subclass Dicotyledoneae including 6 families containing 450 genera and over 3900 species of mainly herbaceous plants with alternate leaves. The mustard family (Cruciferae), which has 350 genera and 2500 species is the largest. It occurs in the North Temperate Zone and is characterized by regular, cruciform flowers and tetradynamous (four longer and two shorter) stamens. The fruit is a silique (elongated pod) or silicle (short, broad pod). The family contains many weeds, some ornamentals, and a considerable

forming part. Here, in place of the moving one or a series of rotary cylindrical filters is Each screen-covered cylinder is mounted in a nd operates partially submerged in the dilute making slurry being supplied to it. As the der revolves, water drains through the screen e interior of the cylinder, and the wet sheet med. This sheet is removed at the top of the der, and may be joined to other wet sheets adjacent cylinders to form a thicker laminated sheet or board. The press section and the are essentially the same as with the four-er.

e factors contributing to paper sheet strength

e bonds, which may be largely but not exclu-ly of the hydrogen bond type, develop as the plasticized, swollen fibers are brought into mate contact during the drainage, pressing, and ng processes.

ther types of forming devices are being studied to some extent used commercially, particu-y on tissues and other lightweight sheets.

receive some type of treatment after forma-n and drying in order to enhance certain de-able characteristics. Of such operations carried on the paper machine, machine calendering is e of the most common. Here the sheet is passed rough the nips formed by a series of steel rolls, e held on top of the other. Surface or tub sizing common, particularly with writing papers.

Machine coating is practiced extensively on book d magazine papers. A coating mix is applied to or both surfaces of the dry sheet by any of everal means. The coating is often clay, contained a high solids aqueous suspension which includes suitable adhesives and other ingredients. The sheet must be redried after the coating application.

Paper products. The products of the paper in-dustry are extremely varied. Throughout the cen-turies paper has retained its traditional uses, but the phenomenal growth of the industry during the past few decades has been due largely to new ap-plications and uses of paper and paper-based ma-terials. These new applications for the most part involve converted products.

Three common machine converting operations were mentioned above. For the most part, however, the more complex converting operations take place subsequent to the paper machine operation, and are called off-machine converting. Supercalendering, embossing, coating, waxing, laminating, impreg-nating, saturating, corrugating, and printing are but a few of the common operations. The cor-rugated paperboard shipping container is in wide-spread use, and represents one of the largest single uses of paper. Multiwall sacks are now used ex-tensively for handling granular materials. Food packaging, with its trend toward the small unit packages, has led to extensive paper utilization, with the paper often coated, waxed, resin-impreg-nated, or combined with other foils and films. Building papers and boards now constitute a siza-ble fraction of the industry output. Although rela-tively small in tonnage, one of the fastest growing segments of the industry is the production of filter papers, for both air and liquid filtration.

Just as traditional uses of paper have been over-shadowed by newer and more diversified applica-tions, so also it appears that some of the traditional materials and manufacturing processes may grad-ually lose their unique position. There is significant interest in the use of glass and other mineral fibers, and in some of the synthetic high polymers, as raw materials for specialty papers. The behavior of these materials often makes desirable different processing techniques. Air-forming, for example, or the use of additive bonding agents, is indicated in some instances. The paper industry is in the midst of a period of growth and evolution, from which it may emerge quite changed from its tradi-tional character. See CELLULOSE; CELL WALLS IN PLANTS; WOOD CHEMICALS. [R.P.WH.]

Bibliography. J. P. Casey, *Pulp and Paper*, 2 vols., 1952; J. N. Stephenson (ed.), *Pulp and Paper Manufacture*, 4 vols., 1950–1955.

Papillomatosis, infectious (rabbit)

A nonfatal disease characterized by large, local-ized, fleshy tumors with horny surfaces, which oc-cur on the neck and shoulders of wild cottontail rabbits. It is caused by a virus, and natural infec-tion occurs by inoculation of the skin surface as the rabbit enters his burrow. The disease is also known as Shope papilloma. It can be transmitted to domestic rabbits by filtrates, but little or no virus can be demonstrated thereafter. The papillomas may change to carcinomas (cancer) in either the wild or the domestic rabbit, but the role of the vi-rus in the change is not clear. See ANIMAL VIRUS; ONCOLOGY; TUMOR VIRUSES. [A.E.M.]

Paprika

A type of pepper (*Capsicum annuum*) with non-pungent flesh, grown for its long red fruit. A mem-ber of the plant order Tubiflorales and of American origin, it is most popular in Hungary and adjacent countries. Seeds are removed from the mature fruit and the flesh is ground to prepare the dry condi-ment commonly referred to as paprika. Production in the United States is limited, California is the only important producing state. See PEPPER; TUBI-FLORALES; VEGETABLE GROWING. [H.J.C.]

Parabola

A member of the class of curves that are intersec-tions of a plane with a cone of revolution (see CONIC SECTION). It is obtained when the cutting plane is parallel to an element of the cone. In ana-lytic geometry, the parabola is defined as the locus of points (in a plane) equally distant from a fixed point F (focus) and a fixed line (directrix) not

vary from 60 to 95%, depending upon the conditions employed. One large use for neutral sulfite semichemical pulp is as the corrugating medium in paperboard containers.

Other fibers. Other sources of cellulose fibers are significant, although of secondary importance. Cotton rags were for many years the principal raw material of the industry They are still employed to a minor extent in bond papers, although cotton-containing papers are today derived principally from cotton linters. Grasses, reeds, and agricultural residues find considerable use, principally in Europe and the Far East. The reuse of waste papers is common, particularly for the lower grades of paper products.

Pulp purification. In many cases, wood pulps obtained from the pulping process are not pure or white enough for subsequent use in papermaking. Hence pulp purification is practiced. This may comprise a number of chemical treatments, of which the most common is bleaching The object of bleaching is to render the pulp whiter without excessive degradation of the cellulose. This is usually accomplished through chemical treatments which make the colored constituents soluble in either neutral or alkaline solution, or which render them colorless.

Almost all bleaching of chemical pulps is carried out with chlorine and chlorine compounds. A typical single-stage bleach employs calcium or sodium hypochlorite Economies in chlorine required, with less cellulose degradation, are realized by using a multistage bleaching sequence, probably using elemental chlorine in the first stage, followed by a caustic-extraction stage (for removal of alkali-soluble compounds) and one or more hypochlorite stages. For ultrahigh-brightness pulps, chlorine dioxide may be used in a final bleaching stage

Other bleaching agents, notably hydrogen and sodium peroxide, sulfur dioxide, and sodium and zinc hydrosulfite are employed, particularly with groundwood and semichemical pulps

Paper manufacture. Paper manufacture is largely a mechanical operation, although the chemical and physicochemical aspects are all-important in determining the final sheet properties. The tendency of cellulose fibers to bond together, when dried from a water suspension, provides the essence of papermaking technology.

Stock preparation. Unmodified cellulose fibers, as obtained from pulping and bleaching operations, are generally unsuited for papermaking They must first be refined; the refining operation is conducted mechanically in beaters or refiners

During refining, the pulp fibers are separated, crushed, frayed, fibrillated, and cut. They imbibe water and swell, becoming more flexible and more pliable Their capacity to bond with one another on drying is greatly enhanced, partially through modification of the fiber surfaces and partially because of the creation of new surface area. Papers made from lightly beaten stocks are typically of low density, soft, and porous, whereas papers from highly beaten stocks are dense, hard, and [illegible] stronger. With given pulps, final paper prop[illegible] are largely controlled through the type and e[illegible] of refining action employed

A variety of additive materials are introduc[illegible] the papermaking pulps during stock prepar[illegible] Fillers, such as clays, titanium dioxide, or cal[illegible] carbonate, are used for the control of sheet op[illegible] and for other secondary reasons. Many [illegible] grades are sized to impart a degree of [illegible] resistance, rosin, usually introduced as a par[illegible] saponified solution-suspension and then pre[illegible] tated on the fibers with alum, is the most [illegible] monly used sizing agent. Dyestuffs are use[illegible] tensively for color control, even with white sh[illegible] Other additives, such as wet-strength agents [illegible] flocculating agents, and defoamers, are use[illegible] needed.

The operations of stock preparation have t[illegible] tionally been carried out in a Hollander be[illegible] This consists of an elongated tub with a ce[illegible] midfeather, and a cylindrical beater roll opera[illegible] against a bedplate on one side of the midfea[illegible] Both the roll and the bedplate have bars on [illegible] surfaces, the clearance between these two set[illegible] bars being adjustable The operation is batch, [illegible] circulation of the furnish caused by rotation of [illegible] beater roll, and beating of the fibers taking p[illegible] as they pass between the two sets of bars [illegible] tinuous refiners, of both the conical and the [illegible] types, are now used extensively.

Sheet formation and drying. The continu[illegible] paper machine converts a very dilute aque[illegible] suspension of fibers and other ingredients int[illegible] dry sheet of paper at speeds which may vary f[illegible] a few feet to over ½ mile per minute. The mach[illegible] may be one city block in length; even though [illegible] operations are complex, the principles involved [illegible] relatively few and straightforward.

The fourdrinier machine is one of the two ma[illegible] sheetforming devices in widespread use. It c[illegible] sists in essence of a continuously moving wire [illegible] or screen, to which the dilute papermaking slu[illegible] is fed and from which the wet, formed sheet is [illegible] moved continuously. The slurry issues through [illegible] slice onto the wire, as a jet of uniform and pro[illegible] thickness, speed, and consistency As the slu[illegible] travels down the machine on its supporting wi[illegible] it passes over table rolls, suction boxes, and fina[illegible] the suction couch roll Each of these devices cau[illegible] water to drain through the wire, and as the water [illegible] removed, the sheet is formed.

The wet sheet at about 20% consistency is [illegible] moved from the wire at the couch roll and tra[illegible] ferred to a supporting woolen felt, which carri[illegible] it through a series of press rolls. Here more wat[illegible] is expressed mechanically, and the sheet progress[illegible] to the drier section at about 33% consistency. T[illegible] remaining water is removed by evaporation as t[illegible] sheet passes over a series of steam heated cylind[illegible] cal driers, which expose alternate sides of the she[illegible] to the hot dryer surfaces.

The second major type of paper machine, t[illegible] cylinder machine, differs from the fourdrinier on[illegible]

a single point called the focus, and rays e focus are reflected parallel to the axis. LYTIC GEOMETRY; MIRROR OPTICS; PARAB-LOIDIC SURFACE; SURFACE AND SOLID OF TION [J.S F.]

ior

lated value which is a function of the sur-sion, molecular weight, and density of a and which is closely related to molecular e. It is defined by the equation

$$P = M\gamma^{1/4}/(d_l - d_v)$$

expression, M is the molecular weight, γ is

res below the critical, d_v is very small com-to d_l, and is usually omitted in the evaluation parachor.

eoretical basis for the parachor has not been stablished Its formulation by S. Sugden in was based upon experimental evidence out-earlier by D B MacLeod. This evidence d that for a given liquid, the ratio of the root of the surface tension to the difference id and vapor densities,

$$\gamma^{1/4}/(d_l - d_v) = \text{constant}$$

ssentially constant over very wide ranges of rature The ratio $M/(d_l - d_v)$ is, except for ery small d_v term, a volume occupied by 1 -molecular weight of the liquid The magni-of the surface tension of a liquid at a given erature is a measure of the forces between molecules of the liquid at that temperature. If, parachors of *two* liquids are evaluated under itions of equal surface tensions, their ratio is of their molar volumes under conditions of lar intermolecular forces This comparison of ir volumes, when it can be made, is often more il than one at a particular temperature and sure See MOLAR VOLUME.

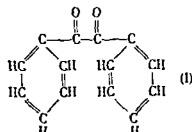
ach atom, group of atoms, closed ring, or multi-bond comprising the molecules of a liquid ributes very nearly independently to its para-r It is, therefore, an additive and constitutive perty For this reason, the parachor is useful determining the structural units comprising the lecules of a liquid.

A number of tabulations of atomic and structural rachors have been presented. These values are based from a correlation of a large number of perimental parachors. The earliest is that by gden in 1924. Several values from this tabulation e listed below.

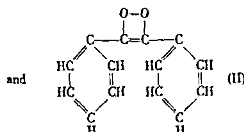
Carbon	4.8	Double bond	23.2
Hydrogen	17.1	Triple bond	46 6
Nitrogen	12.5	4-Membered ring	11 8
Phosphorus	37.7	5-Membered ring	8 5
Oxygen	20 0	6-Membered ring	6.1

More recent tabulations suggest somewhat different values, and include corrections for positions of atoms or groups in a molecule

An example of the use of the parachor in problems of molecular structure concerns the compound benzil, $(C_6H_5)_2C_2O_2$. There are two possible structures consistent with this formula:



and



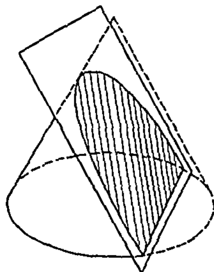
Parachor for (I) is calculated to be 476 0, and for (II) to be 464 4 The experimental parachor is 480 8, suggesting that (I) is correct *See* MOLECULAR STRUCTURE AND SPECTRA; REFRACTION (MOLAR), SURFACE TENSION. [F.J.J.]

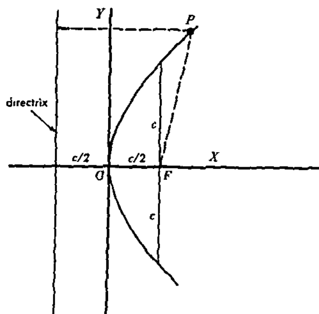
Bibliography: A Weissberger (ed), *Technique of Organic Chemistry*, vol 1, pt. 1, 2d ed., 1949.

Parachute

A flexible, lightweight structure, generally intended to retard the passage of an object within or through atmosphere by materially increasing the resistive surface A parachute is a decelerator or air-braking device in the general form of an oblate hemisphere. It comprises a canopy and cords, which form the suspension and attachment between canopy and object The theory of parachute devices can be traced back to Leonardo Da Vinci; practical application dates from late in the eighteenth century. The parachute, first employed for exhibition purposes, was used as a life-saving vehicle starting with the first years of World War I. and by 1916 was adopted for widespread military use.

Characteristics. A parachute canopy is a membrane which relies upon pressure differential across it to maintain its inflated shape. The differential is created by entrapment of an air mass on the inside and a moving mass of air on the outside Inasmuch as its foremost purpose is to resist the force propelling any body that is to be decelerated, and because that force is most often gravity, the decelerator itself should be extremely light. Rules for determining stresses in thin-walled pressure vessels apply to the parachute canopy, which, by the nature of its shape and loading, experiences tension forces only. A parachute is usually composed of one basic element—fiber When fiber is converted





Parabola (a) As a conic section. (b) As a locus of points.

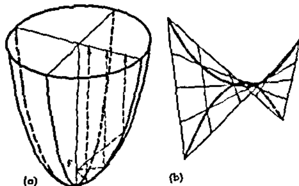
through the point. It is symmetric about the line through F perpendicular to the directrix. To construct a parabola, pin one end of a piece of string to a point F, and fasten the other end to one end of a ruler whose length equals that of the string If the other end of the ruler slides along a line, and the string is kept taut by a pencil, the point of the pencil will trace an arc of a parabola. *Hippocrates of Chios* (ca. 430 B.C.) showed that one of the three famous problems of antiquity, duplication of the cube, can be solved by use of parabolas. The problem is to construct the edge of a cube whose volume is twice *that of a given cube If* c denotes the edge of the given cube, then the desired edge is obtained by considering the two parabolas whose equations in rectangular cartesian coordinates are $x^2 = cy$, $y^2 = 2cx$. They intersect at the origin and a point $P(x_0,y_0)$, with $x_0^3 = 2c^3$.

All parabolas are similar; they differ only scale. For a discussion of the optical propert parabolas, *see* ANALYTIC GEOMETRY. The curve h numerous other properties of interest in both pu and applied mathematics (for example, the traje tory of an artillery shell, assumed to be acted up only by the force of gravity, is a parabola; a the circle that circumscribes the triangle formed any three tangents of a parabola goes through t focus). Archimedes found the area bounded by arc of a parabola and its chord; for example, t area bounded by the parabola $y^2 = 2cx$ and i latus rectum (the chord through F perpendicul to the axis) is $(4\sqrt{2/3})c^2$. The length of the par bolic arc cut off by the latus rectum is $[\sqrt{2}+$ $(1+\sqrt{2})]c$ The volume obtained by revolving th about the parabola's axis is πc^3, and the surface $2/3(3\sqrt{3}-1)\pi c^2$. *See* PARABOLOID. [L.M.BL

Paraboloid

A quadric surface (a surface having an equation second degree in three variables) which has an ax such that every section of the surface by a plan parallel to the axis is a parabola or straight lin *whereas every section by a plane perpendicular* the axis is a central conic with center on the axi If these central conics are all ellipses (or circles a single point), the paraboloid is called an ellipti paraboloid; and in particular if the ellipses are ci cles, the paraboloid is a paraboloid of revolutio Otherwise, these sections normal to the axis a hyperbolas (or a pair of lines through the axis) and the paraboloid is a hyperbolic paraboloid. the z axis is chosen as the axis of the paraboloi the equations may be reduced to the following for the elliptic paraboloid, $z = (x/a)^2 + (y/b)^2$ and for the hyperbolic paraboloid, $z = (x/a)^2 -$ $(y/b)^2$, or $z = cxy$. The hyperbolic paraboloid i a ruled surface. The tangent plane at any point o the surface cuts the surface in two lines, called rul ings For the surface $z = (x/a)^2 - (y/b)^2$, one o these rulings is parallel to (or lies in) the plan $x/a = y/b$, and the other is parallel to (or lies in the plane $x/a = -y/b$.

Paraboloids of revolution are used as reflectin surfaces in automobile headlights and reflectin telescopes. *Rays parallel to the axis are all reflecte*



Two types of paraboloid. (a) Paraboloid of revolution (b) Hyperbolic paraboloid.

rough a single point called the focus, and rays from the focus are reflected parallel to the axis. *See* ANALYTIC GEOMETRY, MIRROR OPTICS; PARABOLA, QUADRIC SURFACE; SURFACE AND SOLID OF REVOLUTION [J.S F.]

Parachor

A calculated value which is a function of the surface tension, molecular weight, and density of a liquid, and which is closely related to molecular structure It is defined by the equation

$$P = M\gamma^{1/4}/(d_l - d_v)$$

In this expression, M is the molecular weight, γ is the surface tension in dynes/cm, d_l is the density of the liquid in g/cm³, and d_v is the density in g/cm³ of the vapor in equilibrium with it. At temperatures below the critical, d_v is very small compared to d_l, and is usually omitted in the evaluation of the parachor.

A theoretical basis for the parachor has not been well established. Its formulation by S. Sugden in 1924 was based upon experimental evidence outlined earlier by D B MacLeod. This evidence showed that for a given liquid, the ratio of the fourth root of the surface tension to the difference of liquid and vapor densities,

$$\gamma^{1/4}/(d_l - d_v) = \text{constant}$$

was essentially constant over very wide ranges of temperature. The ratio $M/(d_l - d_v)$ is, except for the very small d_v term, a volume occupied by 1 gram molecular weight of the liquid The magnitude of the surface tension of a liquid at a given temperature is a measure of the forces between the molecules of the liquid at that temperature If, then, parachors of two liquids are evaluated under conditions of equal surface tensions, their ratio is that of their molar volumes under conditions of similar intermolecular forces. This comparison of molar volumes, when it can be made, is often more useful than one at a particular temperature and pressure *See* MOLAR VOLUME.

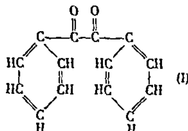
Each atom, group of atoms, closed ring, or multiple bond comprising the molecules of a liquid contributes very nearly independently to its parachor It is, therefore, an additive and constitutive property For this reason, the parachor is useful in determining the structural units comprising the molecules of a liquid

A number of tabulations of atomic and structural parachors have been presented These values are obtained from a correlation of a large number of experimental parachors The earliest is that by Sugden in 1924 Several values from this tabulation are listed below

Carbon	4 8	Double bond	23 2
Hydrogen	17.1	Triple bond	46 6
Nitrogen	12.5	4-Membered ring	11 8
Phosphorus	37.7	5 Membered ring	8.5
Oxygen	20 0	6-Membered ring	6.1

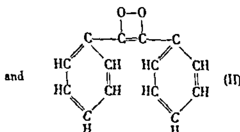
More recent tabulations suggest somewhat different values, and include corrections for positions of atoms or groups in a molecule.

An example of the use of the parachor in problems of molecular structure concerns the compound benzil, $(C_6H_5)_2C_2O_2$ There are two possible structures consistent with this formula:



(I)

and



(II)

Parachor for (I) is calculated to be 476.0, and for (II) to be 464 4. The experimental parachor is 480 8, suggesting that (I) is correct. *See* MOLECULAR STRUCTURE AND SPECTRA, REFRACTION (MOLAR), SURFACE TENSION [F.J.J.]

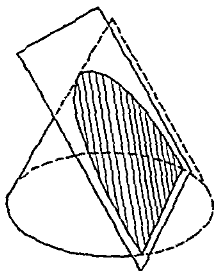
Bibliography· A. Weissberger (ed), *Technique of Organic Chemistry*, vol 1, pt 1, 2d ed., 1949.

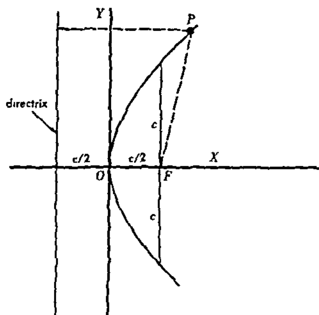
Parachute

A flexible, lightweight structure, generally intended to retard the passage of an object within or through atmosphere by materially increasing the resistive surface. A parachute is a decelerator or air-braking device in the general form of an oblate hemisphere. It comprises a canopy and cords, which form the suspension and attachment between canopy and object. The theory of parachute devices can be traced back to Leonardo Da Vinci; practical application dates from late in the eighteenth century. The parachute, first employed for exhibition purposes, was used as a life-saving vehicle starting with the first years of World War I, and by 1916 was adopted for [illegible]

[illegible]

[illegible] as its foremost purpose is to resist the force propelling any body that is to be decelerated, and because that force is most often gravity, the decelerator itself should be extremely light Rules for determining stresses in thin-walled pressure vessels apply to the parachute canopy, which, by the nature of its shape and loading, experiences tension *forces only* A parachute is usually composed of one basic element—fiber When fiber is converted





Parabola. (a) As a conic section. (b) As a locus of points.

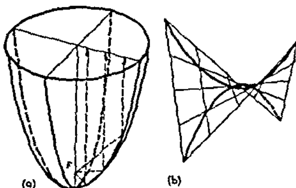
through the point. It is symmetric about the line through F perpendicular to the directrix. To construct a parabola, pin one end of a piece of string to a point F, and fasten the other end to one end of a ruler whose length equals that of the string. If the other end of the ruler slides along a line, and the string is kept taut by a pencil, the point of the pencil will trace an arc of a parabola Hippocrates of Chios (ca. 430 B.C.) showed that one of the three famous problems of antiquity, duplication of the cube, can be solved by use of parabolas. The problem is to construct the edge of a cube whose volume is twice that of a given cube If c denotes the edge of the given cube, then the desired edge is obtained by considering the two parabolas whose equations in rectangular cartesian coordinates are $x^2 = cy$, $y^2 = 2cx$. They intersect at the origin and a point $P(x_0, y_0)$, with $x_0^3 = 2c^3$.

All parabolas are similar; they differ only in scale. For a discussion of the optical property of parabolas, *see* ANALYTIC GEOMETRY. The curve has numerous other properties of interest in both pure and applied mathematics (for example, the trajectory of an artillery shell, assumed to be acted upon only by the force of gravity, is a parabola; and the circle that circumscribes the triangle formed by any three tangents of a parabola goes through the focus). Archimedes found the area bounded by an arc of a parabola and its chord; for example, the area bounded by the parabola $y^2 = 2cx$ and its latus rectum (the chord through F perpendicular to the axis) is $(4\sqrt{2/3})c^2$. The length of the parabolic arc cut off by the latus rectum is $[\sqrt{2} + \ln(1 + \sqrt{2})]c$. The volume obtained by revolving this about the parabola's axis is πc^3, and the surface is $2/3(3\sqrt{3} - 1)\pi c^2$. *See* PARABOLOID. [L.M.BL.]

Paraboloid

A quadric surface (a surface having an equation of second degree in three variables) which has an axis such that every section of the surface by a plane parallel to the axis is a parabola or straight line, whereas every section by a plane perpendicular to the axis is a central conic with center on the axis. If these central conics are all ellipses (or circles or a single point), the paraboloid is called an elliptic paraboloid; and in particular if the ellipses are circles, the paraboloid is a paraboloid of revolution. Otherwise, these sections normal to the axis are hyperbolas (or a pair of lines through the axis), and the paraboloid is a hyperbolic paraboloid. If the z axis is chosen as the axis of the paraboloid, the equations may be reduced to the following: for the elliptic paraboloid, $z = (x/a)^2 + (y/b)^2$; and for the hyperbolic paraboloid, $z = (x/a)^2 - (y/b)^2$, or $z = cxy$. The hyperbolic paraboloid is a ruled surface. The tangent plane at any point of the surface cuts the surface in two lines, called rulings. For the surface $z = (x/a)^2 - (y/b)^2$, one of these rulings is parallel to (or lies in) the plane $x/a = y/b$, and the other is parallel to (or lies in) the plane $x/a = -y/b$

Paraboloids of revolution are used as reflecting surfaces in automobile headlights and reflecting telescopes. Rays parallel to the axis are all reflected



Two types of paraboloid. (a) Paraboloid of revolution (b) Hyperbolic paraboloid.

rough a single point called the focus, and rays im the focus are reflected parallel to the axis. *See* ANALYTIC GEOMETRY; MIRROR OPTICS; PARAB-A, QUADRIC SURFACE; SURFACE AND SOLID OF VOLUTION [J.S F.]

arachor

calculated value which is a function of the surface tension, molecular weight, and density of a quid, and which is closely related to molecular tructure. It is defined by the equation

$$P = M\gamma^{1/4}/(d_l - d_v)$$

n this expression, M is the molecular weight, γ is he surface tension in dynes/cm, d_l is the density of the liquid in g/cm³, and d_v is the density in g/cm³ of the vapor in equilibrium with it. At temperatures below the critical, d_v is very small compared to d_l, and is usually omitted in the evaluation of the parachor.

A theoretical basis for the parachor has not been well established. Its formulation by S. Sugden in 1924 was based upon experimental evidence outlined earlier by D. B. MacLeod. This evidence showed that for a given liquid, the ratio of the fourth root of the surface tension to the difference in liquid and vapor densities,

$$\gamma^{1/4}/(d_l - d_v) = \text{constant}$$

was *essentially* constant over very wide ranges of temperature. The *ratio* $M/(d_l - d_v)$ is, except for the very small d_v term, a volume occupied by 1 gram molecular weight of the liquid The magnitude of the surface tension of a liquid at a given temperature is a measure of the forces between the molecules of the liquid at that temperature If, then, parachors of two liquids are evaluated under conditions of *equal* surface tensions, their ratio is that of their *molar* volumes under conditions of similar intermolecular *forces*. This comparison of molar volumes, when it can be made, is often more useful than one at a particular temperature and pressure [illegible]

E[illegible]

ple [illegible]

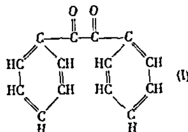
[illegible] molecules of a liquid contributes very *nearly* independently to its parachor It is, therefore, an *additive* and constitutive property For this reason, *the* parachor is useful in determining the structural units comprising the molecules of a liquid

A number of tabulations *of* atomic and structural parachors have been *presented* These values are obtained from a correlation of a large *number* of experimental parachors The earliest is that by Sugden in 1924 Several values *from this* tabulation are listed below

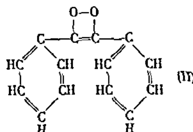
Carbon	4 8	Double bond	23 2
Hydrogen	17.1	Triple bond	46 6
Nitrogen	12.5	4-Membered ring	11 8
Phosphorus	[illegible]	5 Membered ring	8 5
Oxygen	[illegible]	6-Membered ring	6 1

More recent tabulations suggest somewhat different values, and include corrections for positions of atoms or groups in a molecule.

An example of the use of the parachor in problems of molecular structure concerns the compound benzil, $(C_6H_5)_2C_2O_2$. There are two possible structures consistent with this formula:



and



Parachor for (I) is calculated to be 476.0, and for (II) to be 464.4 The experimental parachor is 480.8, suggesting that (I) is correct *See* MOLECULAR STRUCTURE AND SPECTRA, REFRACTION (MOLAR), SURFACE TENSION [F.J.J.]

Bibliography. A Weissberger (ed.), *Technique of Organic Chemistry*, vol. 1, pt. 1, 2d ed., 1949.

Parachute

A flexible, lightweight structure, generally intended to retard the passage of an object within or through atmosphere by materially increasing the resistive surface A parachute is a decelerator or air-braking device in the general form of an oblate hemisphere. It comprises a canopy and cords, which form the suspension and attachment between canopy and object The theory of parachute devices can be traced back to Leonardo Da Vinci; practical application dates from late in the eighteenth century. The parachute, first employed for *exhibition* purposes, was used as a life-saving vehicle starting with the first years of World War I, and by 1916 was adopted for widespread military use

Characteristics. A parachute canopy is *a membrane* which relies upon pressure differential across [illegible]

[illegible] as its foremost purpose is to resist the force propelling any body that is to be decelerated, and because that force is most often gravity, the decelerator itself should be extremely light Rules for determining stresses in thin-walled pressure vessels apply to the parachute canopy, which, by the *nature of its shape* and loading, experiences tension *forces* only A parachute is usually composed of one basic element—fiber. When fiber is converted

to thread, thence to cloth, cord, webbing and tape, parachute construction becomes a matter of structural assembly. Important characteristics of a parachute include porosity (either through the fabric or numerous vents), strength of materials, aerodynamic behavior, dynamic behavior, weight, and ability to deploy freely.

Application. *The classical application of the* parachute, like the closely related airplane, is as a man-carrying apparatus. It is the only suitably demonstrated device for emergency descent from aircraft. In various forms and arrangements, parachutes have been used to (1) deploy paratroops in assault operations; (2) distribute supplies from aircraft; (3) restrict rate of descent of bombs, mines, and flares; (4) drop land and water vehicles; (5) decelerate airplanes, drones, and missiles during landing; and (6) recover space capsules, warheads, drones, and weather recorders. Although applications are limited by available materials (usually nylon, silk, cotton, rayon, or plastic film) and design or fabrication practices, usage expands with the development of new techniques, fabrics, and plastics. [S.E W.]

Bibliography: W. D. Brown, *Parachutes*, 1951; *USAF Parachute Handbook*, Wright Air Develop. Center Tech. Rept 55-265 (ASTIA Document AD 118036), 1956

Paracolon bacilli

Gram-negative, short, rod-shaped bacteria found in the intestine of man and animals. While not considered primarily pathogenic, they cause a problem in the differentiation of pathogenic from nonpathogenic organisms in examination of fecal specimens. The paracolon bacilli lack the ability to utilize lactose immediately, that is, the fermentation of lactose is consistently delayed. Although these microorganisms differ from the *Escherichia* in this respect they were once placed in the *Escherichia*. Bergey's *Manual of Determinative Bacteriology* (1957) now places the paracolon bacilli in another genus, the *Paracolobactrum*, but this is a view not held by some authorities. The delayed fermentation of lactose, sucrose or salicin, and the lack of characteristic *Salmonella* antigenic patterns, also separate these bacteria from the *Salmonellae*. The differential techniques by which the enteric pathogens and nonpathogens are separated depend upon various combinations of characteristic reactions also possessed in part by the paracolon bacilli (*see* CULTURE TECHNIQUE). This complicates the examination of stool specimens for pathogenic organisms. *See* BACTERIOLOGY, MEDICAL; ESCHERICHIA; *IMViC TEST*. [A.J W.]

Paracrinoidea

A class of extinct Pelmatozoa characterized by a theca of numerous plates irregularly arranged, uniserial armlike appendages, and by no clear distinction between adoral and aboral surfaces. Pores arranged in rhombs may occur. Their affinities seem to lie with both cystoids and crinoids. The known forms are of Ordovician age. *See* CRINOIDEA; CYSTIDEA; PELMATOZOA; RHOMBIFERA. [H.B F.]

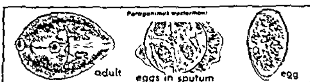
Bibliography: G. Regnell, Non-crinoid pelmatozoa from the Paleozoic of Sweden, *Medd Lunds Geol. Mineral. Inst.*, 108:1-225, 1945.

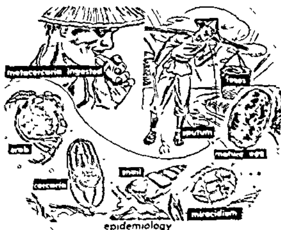
Paraffin

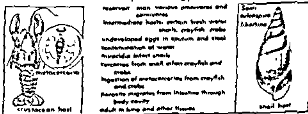
A term used variously to describe either a waxlike substance or a group of compounds. The former use pertains to the high-boiling residue obtained from certain petroleum crudes. It is recovered by freezing out on a cold drum and is purified by crystallization from methyl ethyl ketone. Paraffin wax is a mixture of 26- to 30-carbon alkane hydrocarbons, it melts at 52-57°C. Microcrystalline wax contains compounds of higher molecular weight and has a melting point as high as 90°C. The name paraffin is also used to designate a group of hydrocarbons, although this use is obsolete. These are open-chain compounds of carbon and hydrogen that have only single bonds. Their general formula is C_nH_{2n+2}, where *n* is any integer. *See* ALKANE; WAX, PETROLEUM. [A.L.H.]

Paragonimiasis

Presence of the fluke *Paragonimus westermani* encapsulated in the lungs or other tissues of man. The Orient is the classic endemic focus, but cases of human infection have occurred in America. There are many epidemiologically important reservoir hosts. The disease is connected with eating raw or







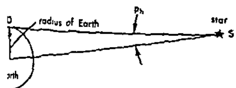
Epidemiology of paragonimiasis. (*From T. T. Mackie, G. W. Hunter, and C. B. Worth, A Manual of Tropical Medicine, 2d ed., Saunders, 1954*)

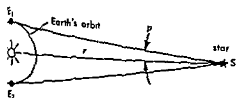
led fresh water crustaceans. Eggs of the fluke passed in sputum or feces, embryonate in water and hatch into the ciliated miracidium which penetrates into snails (*Melania* spp.). Cercariae ... emerge and encyst in the crustacean. When crustacean is ingested by man, the metacercariae migrate by the way of the abdominal cavity ... the diaphragm to the lungs. Ectopic lesions occur. Symptoms resemble those of various diseases; a specific diagnosis is made only by demonstration that the reddish eggs are present. Chemotherapy is not altogether satisfactory. ...DIGENEA; PARASITOLOGY, MEDICAL. [J.F.MA.]

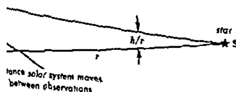
...llax (astronomy)

...difference in direction to a celestial object ... two separated points of observation. The ...nce between observational points is the base. The base line may be provided by the diurnal ...n of Earth ...

...uu, in which case the parallax is annual or ...centric, or it may be provided by the motion ...e solar system within the local galaxy, in ...case the parallax is secular, as illustrated







...e types of parallax. (a) Geocentric or horizontal ... Annual. (c) Secular.

...urnal parallax. The size of Earth produces a ...centric parallax noticeable in observations of ...cts within the solar system. Horizontal parallax ...s defined as the angle between the directions to ...object S at the horizon as seen from the center ...f Earth and from the observer's location O on ...surface of Earth. The mean equatorial hori... ...efers to Earth's *equatorial* radius ... as seen from the mean distance of the celestial object. Thus defined, the diurnal parallax of the Moon is 57′2″.7. Of great historical importance is the measurement of solar parallax; the most accurate measurements give 8″.799 ± ″.001. This value provides the so-called astronomical unit of distance, which is the mean distance from Earth to Sun; it is 149,470,000 ± 17,000 kilometers. The diurnal parallax of stars is negligible because of their great distances.

Annual parallax. The size of Earth's orbit produces a heliocentric parallax noticeable in observations of the nearer stars. This parallax plays a fundamental role in the determination of the distances to stars and is of basic importance in the study of the physical properties of stars. Annual parallax is the maximal angle subtended by one astronomical unit at the star's location, hence the simple relation $p = 1/r$ where p is the annual parallax expressed in seconds of arc, and r the distance expressed in parsecs. One parsec equals 3.26 light years, or 206,265 astronomical units, or 30.84×10^{12} kilometers.

Because of the great distances of the stars, annual parallax of a star was not successfully measured until 1838, accurate determinations became possible in the twentieth century through photography with long-focus refractors (see ASTROMETRY).

Photographic measurement. The observed annual parallactic path of a star on the sky is measured against a background of three or four faint reference stars, presumably at much greater distances, so that, for all practical purposes, they are at infinity. The photographic plates are commonly measured on a long-screw precision measuring engine and are reduced, by conventional algebraic methods, to allow for differences in scale, orientation, and origin of the different plates taken of any one star. The first parallax determinations made in this way were commonly based on some twenty plates, each with two or three exposures, scattered over an interval of two or three years, the observations being centered near maximum parallactic displacement, an accuracy of ±″.01 is thus reached. For the past two decades there has been a tendency to increase the number of plates to reach higher accuracy; however, there is a definite limitation in accuracy, not so much due to the photographic plate and the measuring machine as to systematic errors caused by the telescope and by the atmosphere. The *General Catalogue of Trigonometric Stellar Parallaxes* lists the parallaxes of almost 6000 stars; for the nearer stars the percentage error is often well below 5%. For example, the parallax of Sirius is ″.375 ± ″.004, its distance 2.67 ± .03 parsecs or 8.7 ± .1 light years. On the other hand, the trigonometric parallax method begins to fail for stars beyond a distance of 20 parsecs.

Use of data. Parallax measurements for the hundreds of nearer stars yield basic astronomical information. Knowledge of the spatial properties of the universe is based on accurate geometric measurements made in our immediate cosmic neighbor-

to thread, thence to cloth, cord, webbing and tape, parachute construction becomes a matter of structural assembly. Important characteristics of a parachute include porosity (either through the fabric or numerous vents), strength of materials, aerodynamic behavior, dynamic behavior, weight, and ability to deploy freely.

Application. The classical application of the parachute, like the closely related airplane, is as a man-carrying apparatus. It is the only suitably demonstrated device for emergency descent from aircraft. In various forms and arrangements, parachutes have been used to (1) deploy paratroops in assault operations, (2) distribute supplies from aircraft; (3) restrict rate of descent of bombs, mines, and flares; (4) drop land and water vehicles; (5) decelerate airplanes, drones, and missiles during landing; and (6) recover space capsules, warheads, drones, and weather recorders Although applications are limited by available materials (usually nylon, silk, cotton, rayon, or plastic film) and design or fabrication practices, usage expands with the development of new techniques, fabrics, and plastics. [S.E.W.]

Bibliography: W. D Brown, *Parachutes*, 1951; *USAF Parachute Handbook*, Wright Air Develop Center Tech Rept. 55-265 (ASTIA Document AD 118036), 1956

Paracolon bacilli

Gram-negative, short, rod-shaped bacteria found in the intestine of man and animals While not considered primarily pathogenic, they cause a problem in the differentiation of pathogenic from nonpathogenic organisms in examination of fecal specimens. The paracolon bacilli lack the ability to utilize lactose immediately; that is, the fermentation of lactose is consistently delayed Although these microorganisms differ from the *Escherichia* in this respect they were once placed in the *Escherichia* Bergey's *Manual of Determinative Bacteriology* (1957) now places the paracolon bacilli in another genus, the *Paracolobacterium*, but this is a view not held by some authorities The delayed fermentation of lactose, sucrose or salicin, and the lack of characteristic *Salmonella* antigenic patterns, also separate these bacteria from the *Salmonellae* The differential techniques by which the enteric pathogens and nonpathogens are separated depend upon various combinations of characteristic reactions also possessed in part by the paracolon bacilli (*see* CULTURE TECHNIQUE). This complicates the examination of stool specimens for pathogenic organisms. *See* BACTERIOLOGY, MEDICAL; ESCHERICHIA; IMViC TEST. [A.J.W.]

Paracrinoidea

A class of extinct Pelmatozoa characterized by a theca of numerous plates irregularly arranged, uniserial armlike appendages, and by no clear distinction between adoral and aboral surfaces. Pores arranged in rhombs may occur. Their affinities seem to lie with both cystoids and crinoids. The known forms are of Ordovician age. *See* CRINOIDEA; CYSTIDEA; PELMATOZOA; RHOMBIFE [H.B

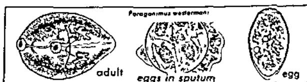
Bibliography: G. Regnell, Non-crinoid pel tozoa from the Paleozoic of Sweden, *Medd. Lu Geol. Mineral. Inst.*, 108:1–225, 1945.

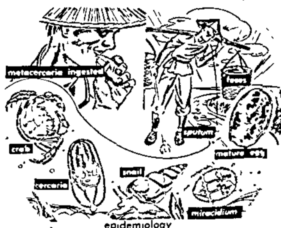
Paraffin

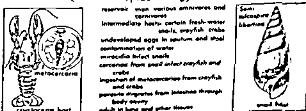
A term used variously to describe either a waxl substance or a group of compounds. The forn use pertains to the high-boiling residue obtair from certain petroleum crudes. It is recovered freezing out on a cold drum and is purified by cr tallization from methyl ethyl ketone. Paraffin w is a mixture of 26- to 30-carbon alkane hydroc bons; it melts at 52–57°C Microcrystalline w contains compounds of higher molecular weight a has a melting point as high as 90°C. The name p affin is also used to designate a group of hydroc bons, although this use is obsolete. These are op chain compounds of carbon and hydrogen that ha only single bonds Their general formula is C_nH_{2n} where *n* is any integer. *See* ALKANE; WAX, PETR LEUM. [A L H

Paragonimiasis

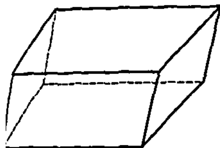
Presence of the fluke *Paragonimus westermani* e capsulated in the lungs or other tissues of ma The Orient is the classic endemic focus, but cas of human infection have occurred in America The are many epidemiologically important reservo hosts. The disease is connected with eating raw c







Epidemiology of paragonimiasis (From T T Mackie, G. W. Hunter, and C. B. Worth, *A Manual of Tropical Medicine*, 2d ed, Saunders, 1954)

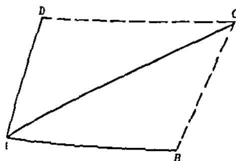


...arallelepiped

...ns that the volume of any parallelepiped is ...al to the product of its base (that is, the area ... face) times the altitude (that is, the distance ... the base to the parallel face). For a rectangu... parallelepiped, the volume is equal to the prod... of the lengths of three edges that meet at a ...tex. See POLYHEDRON, PRISMATOID AND PRIS...M, SOLID (GEOMETRIC). [J.S.F.]

...rallelogram

...quadrilateral, or four-sided polygon, having (a) ...posite sides parallel (definition), or (b) oppo...e sides equal and coplanar, or (c) a pair of op...site sides equal and parallel. Its opposite an...es are equal, its adjacent angles are supplemen-



...ddition of vectors

...ary, and its diagonals bisect each other Its area is the product of its base and altitude The line seg...ments joining the midpoints of consecutive sides of any quadrilateral in space form a parallelogram The parallelogram law for the addition of vectors ...states that if two vectors are represented by the di...rected segments $\overrightarrow{AB}$ and $\overrightarrow{AD}$, their vector sum is represented by the directed diagonal $\overrightarrow{AC}$ of the par...allelogram $ABCD$ (see illustration) See QUADRI...LATERAL, RECTANGLE; SQUARE. [J.S.F.]

Paralysis agitans

A syndrome of the central nervous system caused by diffuse pathology of the brain, resulting in tremor rigidity, and other symptoms.

From an etiological viewpoint, the syndrome may be divided into paralysis agitans (Parkinson's dis...ease), postencephalitic Parkinsonism, and sympto...matic paralysis agitans in combination with liver disease (Wilson's disease) and arteriosclerosis

The pathological lesions of all these disorders are similar and involve degenerative lesions in the basal ganglia of the cerebral hemispheres, and in the cerebral cortex. In postencephalitic Parkinsonism, pathology of the substantia nigra (which is located in the mesencephalon or midbrain) is prominent. The symptoms consist of a coarse tremor of the head and extremities at rest, which diminishes during purposeful movement. Muscular tone is increased; movements are slow and the face acquires a masklike expression. Speech is often slurred. In postencephalitic Parkinsonism, the convergence of the eyes is affected, and patients also exhibit disturbances in their autonomous functions, such as in salivation. There is no mental impairment, although some patients are depressed The differential diagnosis considers chronic alcoholism and hyperthyroidism, and is usually easily made.

The course is progressive, but decades may pass before patients are severely disabled Therapy is symptomatic and includes drugs such as belladonna root and its various components, hyoscine, atropine, strammonium, and scopalamine See ABNORMAL BEHAVIOR. [F.C.R.]

Bibliography· H. H. Merritt, *Textbook of Neurology*, 1955

Paramagnetism

A property exhibited by substances which, when placed in a magnetic field, are magnetized parallel to the field to an extent proportional to the field (except at very low temperatures or in extremely large magnetic fields) Paramagnetic materials always have permeabilities greater than 1, but the values are in general not nearly so great as those of ferromagnetic materials Paramagnetism is of two types, electronic and nuclear

Paramagnetic substances. The following types of substances are paramagnetic.

1 All atoms and molecules which have an odd number of electrons According to quantum mechanics, such a system cannot have a total spin equal to zero, therefore, each atom or molecule has a net magnetic moment which arises from the electron spin angular momentum Examples are organic free radicals and gaseous nitric oxide.

2 All free atoms and ions with unfilled inner electron shells, and many of these ions when in solids or in solution Examples are transition, rare-earth, and actinide elements and many of their salts. This includes ferromagnetic and antiferromagnetic materials above their transition temperatures For a discussion of these materials, *see* ANTIFERROMAGNETISM, FERRIMAGNETISM; FERROMAGNETISM.

3. Several miscellaneous compounds including molecular oxygen and organic biradicals.

4 Metals In this case, the paramagnetism arises from the magnetic moments associated with the spins of the conduction electrons and is called Pauli paramagnetism

Relatively few substances are paramagnetic Aside from the Pauli paramagnetism found in met-

hood Annual parallax is a prerequisite for determining the space velocity of a star and a star's luminosity. The latter, expressed in absolute magnitude M, is related to the apparent magnitude m and the parallax p by $M = m + 5 + 5 \log p$.

In the case of binary stars, parallax is a prerequisite for measuring the total mass $(M_1 + M_2)$ of the binary system (*see* BINARY STARS). The sum of the masses $(M_1 + M_2)$, expressed in terms of the Sun's mass, is related to the space-time dimensions of the double-star orbit through the relation $M_1 + M_2 = (a^3/p^3)(1/P^2)$. Here a is the semimajor axis of the double-star orbit, p the parallax; both a and p are expressed in seconds of arc; P is the binary period expressed in years.

Secular parallax. The motion of the solar system seems attractive for the purpose of measuring stellar distances, because time alone provides an indefinite extension of the base line. However, there are severe limitations because of the motions of the stars themselves; hence only values of the average parallax for groups of stars, called mean secular parallaxes, may be measured. The method has proved important to extend our geometric knowledge of the Milky Way system, in a statistical fashion, by measuring average distances up to several thousand light years for groups of stars. The method also has been important in obtaining information about the average luminosity for stars whose parallaxes are too small to be measured by annual parallax. Closely related to mean secular parallax is the mean parallax of a group of stars obtained from a comparison of their average radial velocity with average values of their proper motions or components thereof.

Distance of remote stars. Geometric measurements of parallax yield the distances of stars From these distances and the apparent magnitudes of the stars, their absolute magnitudes or luminosities are determined Spectral observations show that stars can be grouped into classes of like character (*see* STAR). On the assumption that all stars of like spectral emission have equal intrinsic luminosity or absolute magnitude M, the apparent magnitude m of a remote star can be measured photometrically and the parallax deduced from $5 \log p + 5 = M - m$. This photometric method is limited by loss of light in space due to scattering or absorption. Another means for obtaining absolute magnitudes for use in photometric determination of parallax is from observed periods of cepheid variables (*see* VARIABLE STAR).

The spectroscopic method is calibrated from nearby stars whose geometric parallaxes are known; the method permits, in principle at least, the determination of stellar parallaxes up to any distance, as long as the star's spectrum can be observed The method works better for stars of certain spectral type than for others.

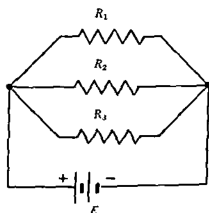
The period-luminosity relation of variable stars has yielded knowledge of the tremendous distances of globular clusters and galaxies because of the appearance of cepheids in those distant systems. The use of this approach, and that of other highly luminous distance indicators, is qualified by their apparently nonunique behavior, as is illustrated by the existence of two types or populations of stars which results in two parallel period-luminosity relations.

Still another method is that of dynamical parallaxes, based on the fact that there is a comparatively small dispersion in stellar masses. From the known apparent size of a double-star orbit and its period of revolution, it is possible—with a reasonable assumption of the sum of the masses—to arrive at a good estimate of the parallax, whose percentage error is only one-third of any percentage error in the assumed combined mass. [P.V.D.K.]

Bibliography: L. F. Jenkins, *General Catalogue of Trigonometric Stellar Parallaxes*, 1952; P. van de Kamp, Elements of long-focus photographic astrometry, *Photogramm. Eng.*, 22(2):32–43, 1956.

Parallel circuit

An electric circuit in which the elements or components are connected between two points with one of the two ends of each component connected to each point. The illustration shows a simple parallel circuit. In more complicated electric networks one or more branches of the network may be made up of various combinations of series or series-parallel elements. *See* CIRCUIT, ELECTRIC.



Parallel circuit

In a parallel circuit the potential difference, or voltage, across each component is the same However, the current through each branch of the parallel circuit may be different For example, the lights and outlets in a house are connected in parallel so that each load will have the same voltage (120 volts), but each load may draw a different current (0.50 amperes in a 60-watt lamp and 10 amperes in a toaster).

For a discussion of parallel circuits *see* ALTERNATING-CURRENT CIRCUIT THEORY, DIRECT-CURRENT CIRCUIT THEORY. [C.T.C.]

Parallelepiped

A polyhedron having six faces that are parallel in pairs Each face is a parallelogram. If adjacent edges are perpendicular, the parallelepiped is called a rectangular parallelepiped—or in common speech, a rectangular box. The formula $V = Bh$

·1) energy levels in a magnetic field (*see* TUM THEORY, NONRELATIVISTIC). A collection h atoms will be distributed among these levels ding to a Boltzmann distribution. The magtion of such a system may be computed by ig the average component of angular momenparallel to the field. The result is

$$M = NgJ\mu_B B_J(a^*)$$

e g is the spectroscopic splitting factor (the ure of the energy level splittings of the sys-, μ_B is the Bohr magneton, $a^* = gJ\mu_B H/kT$, $B_J(a^*)$ is the Brillouin function of a^*

$$_J(a^*) = \frac{2J+1}{2J}\coth\frac{(2J+1)a^*}{2J} - \frac{1}{2J}\coth\frac{a^*}{2J}$$

Brillouin function also enters the theory of omagnetism If a^* is much less than unity, which good approximation except at very low temtures or in large fields, then

$$B_J(a^*) \cong g(J+1)\mu_B H/3kT$$

this case a Curie law again prevails·

$$\chi = M/H = NJ(J+1)g^2\mu_B{}^2/3kT \qquad (4)$$

e effective magneton number is defined by $\overline{J(J+1)}$ and is the quantity usually given in perimental results

If only the electron spin contributes to the total gular momentum, $J = ½$ and $B_{1/2}(a^*) =$ th (a^*) so that, except at low temperatures or gh fields,

$$\chi = N\mu_B{}^2/3kT$$

hich agrees with the classical result. This case is ferred to as the "spin-only" case.

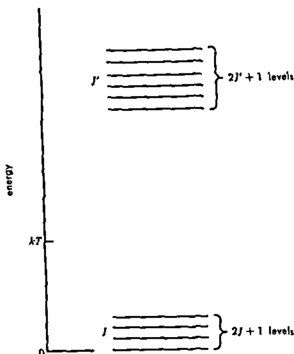
Rare-earth ions. The paramagnetism of rarerth ions at room temperature is summarized by me representative examples in Table 1

The calculated effective magneton numbers in able 1 are the theoretical values for isolated ions The experimental values are derived from Eq (4) sing experimental values of the paramagnetic susceptibility χ There is good agreement for all rareearth ions with the exception of europium and samarium The experimental results of Table 1 refer to the paramagnetic behavior of rare-earth ions in crystals, different salts of the same ion give the same results

The experimental result is therefore that at room temperature, a crystal containing a number of trivalent rare-earth ions has the paramagnetic susceptibility of that number of free trivalent ions

Table 1. Paramagnetism of some trivalent rare-earth ions

Ion	Electron configuration	Effective magneton number	
		Calculated	Experimental
Ce^{3+}	$4f^1 5s^2p^6$	2.54	2.4
Nd^{3+}	$4f^3 5s^2p^6$	3.62	3.5
Sm^{3+}	$4f^5 5s^2p^6$	0.84	1.5
Eu^{3+}	$4f^6 5s^2p^6$	0.00	3.4
Gd^{3+}	$4f^7 5s^2p^6$	7.94	8.0
Yb^{3+}	$4f^{13} 5s^2p^6$	4.54	4.5



Energy levels in Van Vleck paramagnetism

The reason that there is little influence of the crystalline electric fields on the magnetic behavior is that the electrons responsible for the magnetic moments are in the 4f state and therefore occupy an electronic shell lying well inside the ion, a shell that is shielded from outside influence by the 5s and 5p electrons This is in contrast to the behavior of iron-group ions discussed later

At lower temperatures, the influence of the crystalline electric fields on the electrons becomes more important and the behavior of the susceptibility can become quite complex. In this case, the susceptibility depends upon the orientation of the magnetic field with the crystal axes

The behavior of europium and samarium at room temperature is still explainable on the basis of a theory of free ions if the effect of Van Vleck paramagnetism is included

Van Vleck paramagnetism. This arises when the energy states of an atom or ion divide into two groups, those within an energy kT of the ground (lowest energy) state and those which are separated from the ground state by an energy greater than kT Here k is Boltzmann's constant, and T is the absolute temperature. The situation is sketched in the figure The low-lying states give rise to a susceptibility which follows a Curie law. If these lowlying states arise from a single value of the total angular momentum J, as in the figure, then the quantum-mechanical derivation applies [Eq (4)] The high-lying states give rise to a small temperature-independent susceptibility, an effect which is known as Van Vleck paramagnetism. In intermediate cases, such as in the trivalent europium and samarium ions, the upper states are only a little more than kT away from the ground state so that

als, the most important paramagnetic effects are found in the compounds of the transition and rare-earth elements which have partially filled 3d and 4f electron shells respectively.

Electronic paramagnetism. This arises in a substance if its atoms or molecules possess a net electronic magnetic moment. The magnetization arises because of the tendency of a magnetic field to orient the electronic magnetic moments parallel to itself.

The magnitudes of electronic magnetic moments are of the order of a Bohr magneton, which is equal to 9.27×10^{-21} electromagnetic units (erg/gauss). See ELECTRON SPIN.

Nuclear paramagnetism. This arises when there is a net magnetic moment due to the magnetic moments of the nuclei in a substance. An example is solid sodium, in which each sodium atom has a nuclear magnetic moment of 2.217 nuclear magnetons. One nuclear magneton is equal to 5.05×10^{-24} emu Nuclear magnetic moments are about 1000 times smaller than electron magnetic moments. As a result, nuclear paramagnetism produces effects 10^6 times smaller than electron paramagnetic or diamagnetic effects (see DIAMAGNETISM). Therefore, it is usually impossible to detect nuclear paramagnetism by static methods since it will be masked by electronic effects (An exception is the case of nuclear paramagnetism arising from the protons in solid hydrogen) However, paramagnetic effects of nuclei are directly observable in resonance experiments. See MAGNETIC RESONANCE; NUCLEAR MOMENTS.

Langevin theory. The Langevin theory of paramagnetism (P. Langevin, 1905) treats the paramagnetic substance as a classical (non-quantum-mechanical) collection of permanent magnetic dipoles with no interactions between them. The dipoles are the magnetic moments of the paramagnetic atoms or ions in the substance. The first task of a theory of paramagnetism is to account for the experimentally observed susceptibility (ratio of magnetization to applied field) See SUSCEPTIBILITY, MAGNETIC.

If an external magnetic field is applied to the paramagnet, each magnetic dipole experiences a torque. Associated with the force which produces this torque is a potential energy

$$V = -\mu H \cos\theta \qquad (1)$$

where μ is the magnetic moment of the dipole, H is the applied magnetic field intensity, and θ is the angle between the dipole and the direction of H. Now, in the absence of thermal agitation, each permanent magnetic dipole will become oriented in such a way that this potential energy is minimized, that is, oriented parallel to the magnetic field. With all the dipoles lined up, the magnetization (magnetic moment per unit volume), if there are N dipoles per unit volume, would be

$$M = N\mu$$

where the direction of the magnetization would be that of the applied field. Note that in this case an arbitrarily small magnetic field causes all the dipoles to line up so that the susceptibility $\chi = M/H$ would be infinite. In the actual case, there is thermal agitation which in part offsets the aligning tendency of the magnetic field The Langevin theory takes this into account and predicts the paramagnetic susceptibility as a function of temperature.

In the presence of thermal agitation, the magnetic dipoles are not all lined up in the direction of the magnetic field, but there is some distribution of angles made with the field. In this case, the magnetization is

$$M = N\mu\,\overline{\cos\theta} \qquad (2)$$

where $\overline{\cos\theta}$ is the average of the cosine of the angle between dipole and field. The average is taken over the distribution of dipoles in thermal equilibrium. According to statistical mechanics, this average is given by

$$\overline{\cos\theta} = \int e^{(-V/kT)} \cos\theta\, d\Omega / \int e^{(-V/kT)}\, d\Omega$$

where $d\Omega$ is the element of solid angle and $e^{(-V/kT)}$ is the Boltzmann distribution in energy $V = -\mu H \cos\theta$ [see Eq (1)] of a dipole at angle θ with respect to the applied field at absolute temperature T. The integrations may be performed and the result is $L(a)$, the Langevin function of $a = \mu H/kT$ (see LANGEVIN FUNCTION). The result may be combined with Eq. (2) to give

$$M = N\mu L(a)$$

If $a \ll 1$, then $L(a) \cong a/3$ so that

$$M \cong N\mu^2 H/3kT$$

This is a good approximation except at low temperatures or extremely high fields. The susceptibility is

$$\chi = M/H = N\mu^2/3kT = C/T \qquad (3)$$

The $1/T$ dependence of the susceptibility is known as the Curie law. The Curie law was established empirically by P. Curie in 1895 and is obeyed by many gases, liquids, and solids. There are some paramagnetic solids which obey the Curie-Weiss law $\chi = C/(T - \Theta)$ in a certain temperature range. Here Θ is the Curie temperature. The modification often arises because of effective interactions between the dipoles which are neglected in the preceding development It may also be due to distortion effects. See CURIE-WEISS LAW.

Experimental data for the paramagnetic susceptibility is often expressed in terms of the effective magnetic moment which must be used for μ in the Curie law [Eq (3)] in order to give the observed slope of the curve of χ plotted against $1/T$.

Quantum theory. The quantum-mechanical theory of paramagnetism was worked out in detail by J. H. Van Vleck in 1928. This theory is based on the fact that the magnetic moment of the permanent magnetic dipole arises from the total angular momentum of the electrons in the paramagnetic atom, ion, or molecule. Thus an atom with total angular momentum quantum number J has

$(2J+1)$ energy levels in a magnetic field (see QUANTUM THEORY, NONRELATIVISTIC). A collection of such atoms will be distributed among these levels according to a Boltzmann distribution. The magnetization of such a system may be computed by finding the average component of angular momentum parallel to the field The result is

$$M = NgJ\mu_B B_J(a^*)$$

where g is the spectroscopic splitting factor (the measure of the energy level splittings of the system), μ_B is the Bohr magneton, $a^* = gJ\mu_B H/kT$, and $B_J(a^*)$ is the Brillouin function of a^*.

$$B_J(a^*) = \frac{2J+1}{2J}\coth\frac{(2J+1)a^*}{2J} - \frac{1}{2J}\coth\frac{a^*}{2J}$$

The Brillouin function also enters the theory of ferromagnetism If a^* is much less than unity, which is a good approximation except at very low temperatures or in large fields, then

$$B_J(a^*) \cong g(J+1)\mu_B H/3kT$$

In this case a Curie law again prevails:

$$\chi = M/H = NJ(J+1)g^2\mu_B^2/3kT \qquad (4)$$

The effective magneton number is defined by $g\sqrt{J(J+1)}$ and is the quantity usually given in experimental results

If only the electron spin contributes to the total angular momentum, $J = ½$ and $B_{1/2}(a^*) = \tanh(a^*)$ so that, except at low temperatures or high fields,

$$\chi = N\mu_B^2/3kT$$

which agrees with the classical result This case is referred to as the "spin-only" case

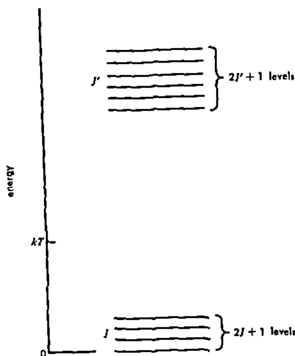
Rare earth ions. The paramagnetism of rare-earth ions at room temperature is summarized by some representative examples in Table 1

The calculated effective magneton numbers in Table 1 are the theoretical values for isolated ions The experimental values are derived from Eq (4) using experimental values of the paramagnetic susceptibility χ There is good agreement for all rare-earth ions with *the* exception of europium and samarium The experimental results of Table 1 refer to the paramagnetic behavior of rare-earth ions in crystals, different salts of the same ion give the same results

The experimental result is therefore that at room temperature, a crystal containing a number of trivalent rare-earth ions has the paramagnetic susceptibility of that *number* of free trivalent ions

Table 1 Paramagnetism of some trivalent rare-earth ions

Ion	Electron configuration	Effective magneton number	
		Calculated	Experimental
Ce^{3+}	$4f^{1}5s^{2}p^{6}$	2.54	2.4
Nd^{3+}	$4f^{3}5s^{2}p^{6}$	3.62	3.5
Sm^{3+}	$4f^{5}5s^{2}p^{6}$	0.84	1.5
Eu^{3+}	$4f^{6}5s^{2}p^{6}$	0.00	3.4
Gd^{3+}	$4f^{7}5s^{2}p^{6}$	7.94	8.0
Yb^{3+}	$4f^{13}5s^{2}p^{6}$	4.54	4.5



Energy levels in Van Vleck paramagnetism

The reason that there is little influence of the crystalline electric fields on the magnetic behavior is that the electrons responsible for the magnetic moments are in the $4f$ state and therefore occupy an electronic shell lying well inside the ion, a shell that is shielded from outside influence by the $5s$ and $5p$ electrons This is in contrast to the behavior of iron-group ions discussed later.

At lower temperatures, the influence of the crystalline electric fields on the electrons becomes more important and the behavior of the susceptibility can become quite complex In this case, the susceptibility depends upon the orientation of the magnetic field with the crystal axes

The behavior of europium and samarium at room temperature is still explainable on the basis of a theory of free ions if the effect of Van Vleck paramagnetism is included.

Van Vleck paramagnetism. This arises when the energy states of an atom or ion divide into two groups, those within an energy kT of the ground (lowest energy) state and those which are separated from the ground state by an energy greater than kT Here k is Boltzmann's constant, and T is the absolute temperature The situation is sketched in the figure The low-lying states give rise to a susceptibility which follows a Curie law If these low-lying states arise from a single value of the total angular momentum J, as in the figure, then the quantum-mechanical derivation applies [Eq. (4)] The high-lying states give rise to a small temperature-independent susceptibility, an effect which is known as Van Vleck paramagnetism In intermediate cases, such as in the trivalent europium and samarium ions, the upper states are only a little more than kT away from the ground state so that

Table 2. Paramagnetism of iron-group ions

Ion	Electron configuration	Effective magneton number: Calculated with J	Calculated with S only	Experimental
Ti^{3+}, V^{4+}	$3d^1$	1.55	1 73	1 8
V^{3+}	$3d^2$	1 63	2 83	2 8
Cr^{3+}, V^{2+}	$3d^3$	0 77	3 87	3 8
Mn^{3+}, Cr^{2+}	$3d^4$	0 00	4 90	4 9
Fe^{3+}, Mn^{2+}	$3d^5$	5 92	5.92	5 9
Fe^{2+}	$3d^6$	6 70	4 90	5.4
Co^{2+}	$3d^7$	6 54	3 87	4 8
Ni^{2+}	$3d^8$	5 59	2 83	3 2
Cu^{2+}	$3d^9$	3 55	1 73	1 9

the temperature dependence is still more complicated.

Iron-group ions. The paramagnetism of iron-group ions in crystals is summarized in Table 2.

Quenching of orbital angular momentum is exhibited in crystals containing ions of the iron group. The last three columns of Table 2 indicate *that the orbital angular momentum makes no contribution* to the magnetic moment but that the iron-group ions behave in crystals as free ions with only the spin S contributing to the magnetic effects. This is evidenced by the fact that the "spin-only" values of the effective magneton numbers agree well with the experimental results. The orbital angular momentum is quenched because the $3d$ electronic shell, which gives rise to the paramagnetism, is outermost for the iron group, it is therefore exposed to the strong crystalline electric fields arising from neighboring ions. These asymmetric electric fields decouple the orbital angular momentum from the spin angular momentum This *means that the energy levels are no longer specified* by the total angular momentum quantum number J, S alone may determine the levels. More precisely, the $(2L+1)$ degenerate orbital angular momentum states of orbital angular momentum quantum number L may be split by the crystal fields so that the lowest orbital state is nondegenerate (singlet). Then there is no possibility of orienting the orbital angular momentum by a magnetic field so that only the spin contributes to the magnetic moment. It is often said that the crystal field "locks" the orbital angular momentum so that it cannot be oriented by a magnetic field. Partial quenching occurs when the orbital degeneracy is only partially removed by the crystal field. Partial quenching and anisotropic effects can also be caused by spin-orbit coupling. This influence may be observed in spin-resonance and specific-heat experiments. *See* ADIABATIC DEMAGNETIZATION; MAGNETIC RESONANCE.

Pauli paramagnetism. This is the paramagnetism associated with the conduction electrons of a metal. A metal is usually described in terms of a collection of positive ions with closed shells which are arranged on a *crystal lattice plus electrons* which are essentially free to move about the crystal (*see* FREE-ELECTRON THEORY OF METALS). Each electron has an intrinsic spin angular momentum and these momenta give rise to a paramagnetic magnetic moment. *At first sight it would seem to be* correct to apply the Langevin formula to this "gas" of electrons, but the experimental facts are that the paramagnetic susceptibility of conduction electrons is about one-hundredth of that predicted by the Langevin formula [Eq. (3)]. Furthermore, the susceptibility is temperature independent rather than *varying as $1/T$ (Curie law). The explanation* was given by W. Pauli in 1927: The electrons obey the quantum statistics of E. Fermi and P. A. M. Dirac rather than the classical statistics which are used in the derivation of the Langevin formula. This means that a given energy state can be occupied at most by two electrons, and their spin angular momenta must be in opposite directions (*see* EXCLUSION PRINCIPLE). As a result, the net angular momentum is zero, even on application of a magnetic field. Thus most of the electrons in a metal contribute in sum no magnetic moment. That is to say, an electron's spin angular momentum may not orient parallel to an applied magnetic field be*cause there is already an electron in that energy* state with its spin parallel to the field. There are, however, a few electrons which are not "paired off," and the spins of these can be oriented by the field These electrons contribute to the susceptibility according to a Curie law, but the number of them is proportional to the temperature The combination of the two temperature dependences leads to a temperature-independent susceptibility, smaller than the prediction of the Langevin formula for N electrons per unit volume because only a fraction of these may contribute. The Pauli susceptibility may be written (for a free electron gas)

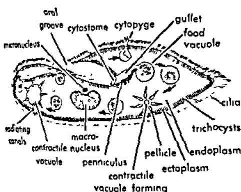
$$\chi = 3N\mu_B^2/2kT_F$$

where N is the number of electrons per unit volume and kT_F is the Fermi energy characteristic of the metal. The fraction of electrons contributing to the susceptibility at temperature T is of the order T/T_F. For sodium, for example, $kT_F = 3.12$ electron volts, $T_F = 37{,}000°K$. The Pauli paramagnetism of metals has been observed in spin-resonance experiments. The total susceptibility arises from Pauli paramagnetism and diamagnetic contributions from conduction electrons and ion cores. [E.A.; F.K.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956; L. Marton (ed.), *Advances in Electronics and Electron Physics*, vol. 6, 1954; J. H Van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, 1932.

Paramecium

A ciliated protozoan of the subphylum Ciliophora. Perhaps the most common of all protozoan genera, there are several species of *Paramecium* which differ in anatomical details They have an oral groove leading to a *gullet through which food* enters. They eat bacteria, smaller protozoa, yeast, and algae (*see* ALGAE; BACTERIA; YEAST). There are two contractile vacuoles for the control of water balance and excretion, and a definite spot for the discharge of solid wastes. Locomotion is ac-



Paramecium caudatum, a fresh-water ciliate. Diagrammatic sketch shows principal parts. (*From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957*)

complished by the beating of the cilia which, as a result of their shape, produce an undulating, spiral movement. In a process called conjugation, the *Paramecium* shows one of the most primitive forms of sexual reproduction. Normal reproduction is by binary fission. *See* CILIOPHORA [J. D. B.]

Parameter

An auxiliary variable, functions of which give the coordinates of a curve or surface. The coordinates of a curve are functions of one parameter. A curve in 3-space has the parametric equations

$$x = f(t) \qquad y = g(t) \qquad z = h(t)$$

The coordinates of a surface are functions of two parameters:

$$x = f(u,v) \qquad y = g(u,v) \qquad z = h(u,v)$$

When constant values are assigned to the parameters u or v, these equations represent the parametric curves on the surface. General surface curves are obtained by replacing u and v by functions of a new parameter t.

An arbitrary constant in an equation is also called a parameter. Variations in the values of the parameter generate a system of equations which may represent a *family* of curves or surfaces. Such families are called one-parameter, two-parameter, and so on, according to the number of independent parameters. *See* PARAMETRIC EQUATION. [L. BR.]

Parametric amplifier

A low noise, high-frequency amplifier that employs nonlinear variation of capacitance or inductance to achieve amplification. The most common type of parametric amplifier uses a back-biased silicon diode, whose capacitance is varied by the application of ultra high or microwave frequency *power*. The parametric amplifier is also known as the variable reactance amplifier, varactor, or MAVAR.

In a device of this type, in which the electrical charges flow through a solid crystalline material, it is possible to obtain current densities many times greater than those possible in an electron beam in vacuum, basically because the electrons do not repel each other in the crystal. Therefore, the dimensions of the device can be kept small, consistent with the requirements of very high-frequency amplification. The parametric diode is intrinsically a low-noise device, because its action is dependent on reactance variation, which introduces no noise of itself, rather than the noise-producing variation of resistance employed in the conventional mixer diode.

Parametric amplifiers require a source of high-frequency power, usually called the pump, rather than a direct-current supply, as for conventional vacuum tubes.

As a low-noise, high-frequency amplifier, the parametric amplifier competes with another solid-state device that also requires a high-frequency pump, the maser. The maser has excellent performance with respect to noise, but it must be refrigerated to temperatures usually less than 4C° above absolute zero, and requires a strong and uniform magnetic field. Compared to the maser, the parametric amplifier is extremely simple. *See* MASER.

Operation. The operation of the parametric amplifier can be most easily explained in terms of its low-frequency analog. The following refers to the capacitive version, but a similar treatment applies to the parametric amplifier using a nonlinear inductor.

Multiplying the cosines of two angular frequencies ω_1 and ω_2

$$\cos\omega_1 \cos\omega_2 = \frac{\cos(\omega_1 + \omega_2)}{2} + \frac{\cos(\omega_1 - \omega_2)}{2}$$

generates terms containing the sum and the difference of the frequencies. Multiplication of two frequency components occurs when they are combined by a nonlinear device, commonly called a mixer. In the parametric diode, nonlinear changes of capacitance with applied voltage are employed to generate the sum and difference components.

The equivalent circuit in its most elementary form consists of two tuned circuits coupled through a nonlinear capacitor. From an external source,

tuned circuit. Through the action of the nonlinear capacitance, multiplication of the ω_s and ω_p components occurs.

It is a property of nonlinear reactive mixing that amplification of signals can occur, with the added signal power being derived from the pump source. Thus amplification of the signal ω_s can be achieved. However, the amplified signal may be at a different frequency ω_i (i stands for idler). The original ω_s signal can be recovered from the ω_i component by standard electronic techniques.

Several different variations are possible. The pump frequency may equal the difference between the signal output frequency and the signal input frequency. This case, sometimes called the non-

Table 2. Paramagnetism of iron-group ions

		Effective magneton number		
Ion	Electron configuration	Calculated with J	Calculated with S only	Experimental
Ti^{3+}, V^{4+}	$3d^1$	1.55	1.73	1.8
V^{3+}	$3d^2$	1.63	2.83	2.8
Cr^{3+}, V^{2+}	$3d^3$	0.77	3.87	3.8
Mn^{3+}, Cr^{2+}	$3d^4$	0.00	4.90	4.9
Fe^{3+}, Mn^{2+}	$3d^5$	5.92	5.92	5.9
Fe^{2+}	$3d^6$	6.70	4.90	5.4
Co^{2+}	$3d^7$	6.54	3.87	4.8
Ni^{2+}	$3d^8$	5.59	2.83	3.2
Cu^{2+}	$3d^9$	3.55	1.73	1.9

the temperature dependence is still more complicated.

Iron-group ions. The paramagnetism of iron-group ions in crystals is summarized in Table 2.

Quenching of orbital angular momentum is exhibited in crystals containing ions of the iron group. The last three columns of Table 2 indicate that the orbital angular momentum makes no contribution to the magnetic moment but that the iron-group ions behave in crystals as free ions with only the spin S contributing to the magnetic effects. This is evidenced by the fact that the "spin-only" values of the effective magneton numbers agree well with the experimental results. The orbital angular momentum is quenched because the $3d$ electronic shell, which gives rise to the paramagnetism, is outermost for the iron group; it is therefore exposed to the strong crystalline electric fields arising from neighboring ions. These asymmetric electric fields decouple the orbital angular momentum from the spin angular momentum This means that the energy levels are no longer specified by the total angular momentum quantum number J; S alone may determine the levels More precisely, the $(2L+1)$ degenerate orbital angular momentum states of orbital angular momentum quantum number L may be split by the crystal fields so that the lowest orbital state is nondegenerate (singlet). Then there is no possibility of orienting the orbital angular momentum by a magnetic field so that only the spin contributes to the magnetic moment. It is often said that the crystal field "locks" the orbital angular momentum so that it cannot be oriented by a magnetic field Partial quenching occurs when the orbital degeneracy is only partially removed by the crystal field Partial quenching and anisotropic effects can also be caused by spin-orbit coupling This influence may be observed in spin-resonance and specific-heat experiments. *See* ADIABATIC DEMAGNETIZATION; MAGNETIC RESONANCE.

Pauli paramagnetism. This is the paramagnetism associated with the conduction electrons of a metal. A metal is usually described in terms of a collection of positive ions with closed shells which are arranged on a crystal lattice plus electrons which are essentially free to move about the crystal (*see* FREE-ELECTRON THEORY OF METALS). Each electron has an intrinsic spin angular momentum and these momenta give rise to a paramagnetic magnetic moment. At first sight it would seem to be correct to apply the Langevin formula to this "gas" of electrons, but the experimental facts are that the paramagnetic susceptibility of conduction electrons is about one-hundredth of that predicted by the Langevin formula [Eq. (3)]. Furthermore, the susceptibility is temperature independent rather than varying as $1/T$ (Curie law). The explanation was given by W. Pauli in 1927: The electrons obey the quantum statistics of E. Fermi and P. A. M. Dirac rather than the classical statistics which are used in the derivation of the Langevin formula This means that a given energy state can be occupied at most by two electrons, and their spin angular momenta must be in opposite directions (*see* EXCLUSION PRINCIPLE). As a result, the net angular momentum is zero, even on application of a magnetic field. Thus most of the electrons in a metal contribute in sum no magnetic moment. That is to say, an electron's spin angular momentum may not orient parallel to an applied magnetic field because there is already an electron in that energy state with its spin parallel to the field There are, however, a few electrons which are not "paired off," and the spins of these can be oriented by the field. These electrons contribute to the susceptibility according to a Curie law, but the number of them is proportional to the temperature. The combination of the two temperature dependences leads to a temperature-independent susceptibility, smaller than the prediction of the Langevin formula for N electrons per unit volume because only a fraction of these may contribute. The Pauli susceptibility may be written (for a free electron gas)

$$\chi = 3N\mu_B^2/2kT_F$$

where N is the number of electrons per unit volume and kT_F is the Fermi energy characteristic of the metal. The fraction of electrons contributing to the susceptibility at temperature T is of the order T/T_F. For sodium, for example, $kT_F = 3.12$ electron volts, $T_F = 37{,}000°K$. The Pauli paramagnetism of metals has been observed in spin-resonance experiments. The total susceptibility arises from Pauli paramagnetism and diamagnetic contributions from conduction electrons and ion cores. [E.A.; F.K.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956, L. Marton (ed.), *Ad-* [illegible] Physics vol 6, [illegible] *ric and*

Paramecium

A ciliated protozoan of the subphylum Ciliophora Perhaps the most common of all protozoan genera, there are several species of *Paramecium* which differ in anatomical details. They have an oral groove leading to a gullet through which food enters. They eat bacteria, smaller protozoa, yeast, and algae (*see* ALGAE; BACTERIA; YEAST). There are two contractile vacuoles for the control of water balance and excretion, and a definite spot for the discharge of solid wastes. *Locomotion is ac-*

of increasing s The parametric formula

$$K = \pm\frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}$$

$+$ or $-$ sign is to be chosen according as ositive or negative Here

$$x'' = \frac{d^2x}{dt^2} \quad y' = \frac{dy}{dt} \quad y'' = \frac{d^2y}{dt^2}$$

$F(x)$ is obtained from $x = f(t), y = g(t)$, by the technique of differentials that $x) \, dx$ and $dx = f'(t) \, dt, \; dy = g'(t) \, dt$,

$$F'(x) = \frac{g'(t)}{f'(t)}$$

$$\frac{d^2y}{dx^2} = \frac{dF'(x)}{dx} = \frac{f'(t)g''(t) - g'(t)f''(t)}{[f'(t)]^3}$$

lerivatives of F can also be found in terms tives of f and g

ne cases it is necessary to include what $t \to \pm\infty$ to complete a parametric repre- in a natural way. For instance,

$$x = \frac{2t}{1+t^2} \quad y = \frac{1-t^2}{1+t^2}$$

its the circle $x^2 + y^2 = 1$ except for the $(0,-1)$ This point is obtained in the limit as $\to -\infty$ or $t \to +\infty$

cycloid $x = a(t - \sin t), \; y = a(1 - \cos t)$ ample of a curve that is easily and naturally nted in parametric form, but is represented th great awkwardness by a single equation y

es in space of three dimensions are repre- parametrically in the form $x = f(t), \; y =$ $z = h(t)$ The conditions for a smooth arc ilar to the conditions in the plane case

metric surface. A parametric surface in of three dimensions is defined by $x = f(u,v)$, $(u,v), \; z = h(u,v)$, where f, g, h are contin- unctions of the

nd. that, for (u,v) in a certain square (or region) in the uv plane, the functions f, g, h continuous first partial derivatives such that the jacobian determinants

$$\begin{vmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \\ \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{vmatrix} \quad J_2 = \begin{vmatrix} \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \\ \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \end{vmatrix} \quad J_3 = \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix}$$

ever zero at the same point (u,v) Under these tions a sufficiently small square containing a point in the uv plane is "mapped" onto a piece of surface, and (J_1, J_2, J_3) are com- nts of a vector which is perpendicular to the tangent to the surface The area A of this of surface is

$$A = \iint \sqrt{J_1^2 + J_2^2 + J_3^2} \, du \, dv$$

where the double integral is extended over the square in the uv plane. For a discussion of implicit functions and jacobian determinants, *see* PARTIAL DIFFERENTIATION *See also* ANALYTIC GEOMETRY; CALCULUS, DIFFERENTIAL AND INTEGRAL; OPERATOR THEORY. [A E.T.]

Paranoia

A mental disease characterized by logically systematized delusions of persecution It is a severe illness and usually takes a chronic course. Pure forms of paranoia are rare. Paranoiacs belong to the small group of mental patients who can be potentially dangerous The thinking and feeling of these patients, apart from their rigid and unalterable delusions, are unimpaired

The chief symptoms are systematic, consistent, and persistent delusions of persecution, far exceeding character traits of hypersensitivity, suspicions and querulousness, irascibility and pettiness which are found in paranoid personalities The paranoiacs, like paranoid personalities, are problems in the community, but to a greater degree. Their hostility, seclusiveness, and cynicism make contact with them difficult

Although the cause of paranoia is not clearly established, psychological factors play a very important role Sigmund Freud drew attention to the etiological significance of repressed and projected homosexuality and aggression The precipitating cause is often an actual tort or injustice inflicted upon the patient Normal persons overcome such an experience, but the paranoiac does not

The differential diagnosis is usually not difficult. Although the Viennese school of psychiatry includes paranoia in the schizophrenic group, the disorder needs to be differentiated from paranoid schizophrenia, paranoia in alcoholism, paranoid states, and paranoid psychopathology *See* PARANOID STATE, SCHIZOPHRENIA.

Hospitalization, in many cases, is an absolute necessity The only method of treatment is psychotherapy which has proven in most cases extremely difficult and not successful *See* PSYCHOSIS; PSYCHOTHERAPY [F.C R.]

Bibliography: J R. Ewalt, E A Strecker, and F G Ebaugh, *Practical Clinical Psychiatry*, 8th ed , 1957

Paranoid state

A severe and chronic psychosis with predominant delusions of persecution The delusions, in contrast to those of paranoia, are inconsistent, illogical, and often accompanied by hallucinations. Otherwise, the personality does not show the disintegration which occurs in schizophrenia However, some schools of psychiatry do not separate this mental disorder from the schizophrenias (*see* PARANOIA; SCHIZOPHRENIA).

The symptoms are an incoherent and bizarre system of delusions of persecution and grandeur, fed by various hallucinatory experiences Patients are

regenerative up-converter, is unconditionally stable for all source and load impedances. Its power gain can be as great as the ratio of output frequency to input frequency

In the remaining cases, the pump frequency is the highest frequency present. In all these, the admittance of the variable-reactance amplifier possesses a negative real part, and the amplifier is therefore regenerative and may become unstable. The regenerative up-converter has a signal output frequency higher than the input signal frequency, with the pump frequency equaling the sum of the input and output frequencies. This connection generally gives highest usable gain and lowest noise figure. The regenerative down-converter has an output frequency lower than the input frequency, with the sum of the input and output frequencies again equal to the pump frequency. The performance of this configuration is somewhat less favorable than the regenerative up-converter, but it has the advantage that the pump frequency is only slightly higher than the input frequency, and that the output frequency is lower than the input frequency.

In the final variation, use is made of the negative resistance presented by the variable-reactance amplifier. The input frequency and output frequency are equal, and it is necessary to resort to a nonreciprocal element, such as a ferrite isolator or circulator, to obtain the useful amplification. For this circuit to operate, an idler circuit, tuned to the frequency difference between the pump and the input signal, must be used.

The inductive reactance form of parametric amplifier may use a ferrite or a garnet as the nonlinear element

Characteristics. Compared to high-frequency vacuum-tube amplifiers, the parametric amplifier has the advantage of lower noise In a single stage, its gain-bandwidth product may be somewhat less than that of the equivalent vacuum-tube stage However, a number of parametric amplifiers are used in tandem in a structure known as a traveling-wave parametric amplifier, to produce extraordinary bandwidth. For a traveling-wave parametric amplifier bandwidth of several hundred megacycles, gains of the order of 20 decibels (db) and a noise figure as low as 0.5 db are theoretically possible. Similar performance can be expected from other types but with reduced bandwidth Present performance does not reach these limits, particularly in noise figure

It is possible to use the properties of electron beams, instead of current in solid-state materials, to produce the variable energy storage needed for parametric amplification. Such electron beam devices have achieved results, in spite of the presence of thermal noise in the electron beam, comparable in noise figure to those obtained with solid-state reactance amplifiers. *See* SEMICONDUCTOR.

[C.V.B.]

Bibliography: W. E. Danielson, Low noise in solid state parametric amplifiers at microwave frequencies, *J. Appl. Phys.*, 30:8–15, 1959.

Parametric equation

A type of mathematical equation used, typically, t represent curves in a plane or in space of three d mensions. In principle, however, there is no limit tion to any particular number of dimensions. A p rameter is actually an independent variable. In el mentary analytic geometry a curve in the xy plar is often studied, in the first instance, as the locus (an equation $y = F(x)$ or $G(x,y) = 0$. The form $y =$ $F(x)$ is not adequate for the complete represent tion of certain curves, whereas the form $G(x,y) =$ 0 may be adequate. The circle $x^2 + y^2 - 16 =$ affords an example. But the form $G(x,y) = 0$ is n always convenient. The parametric form $x = f(t)$ $y = g(t)$ is often the most convenient; moreover, is often the naturally occurring form of represent tion of the curve. For the circle $x^2 + y^2 - 16 =$ one possible parametric representation is

$$x = 4 \cos t, \qquad y = 4 \sin t$$

Parametric curves. A pair of equations $x =$ $f(t)$, $y = g(t)$, where f and g are continuous func tions defined for some interval of values of t, fo example, $a \leq t \leq b$, is said to define a parametri curve. The locus of points (x,y) obtained in thi way is not always what would be regarded as a curve by a layman. If one thinks of t as time, th equations define the motion of the point (x,y) as increases from a to b Clearly the path can cros itself, double back on itself, or the point may ever remain motionless. Even more surprising, the poin may pass through every position within an entir square unit of area.

In calculus the customary restrictions placed or f and g in the case of a parametric curve are that f and g have derivatives f', g' which are continuous and such that $f'(t)$ and $g'(t)$ are not zero at the same time. Exceptions for isolated values of t are sometimes permitted. The result of these conditions is that a part of the curve near any one point on it can be represented either in the form $y = F(x)$ or in the form $x = F(y)$, where F is a differentiable function with a continuous derivative (a result of implicit-function theory). This implies that the parametric curve really looks like a curve in the intuitive sense and has a continuously turning tangent. The curve may cross itself, but the part generated by small variations of t from any fixed value t_0 is what is called a smooth arc.

The arc length L of a smooth arc from t_0 to t_1 is

$$L = \int_{t_0}^{t_1} \left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right]^{1/2} dt$$

The corresponding differential formula is

$$ds^2 = dx^2 + dy^2$$

where s is arc length measured along the curve. In *polar coordinates* the formula is

$$ds^2 = dr^2 + r^2\, d\theta^2$$

The curvature of a curve is defined as $K = d\phi/ds$, where s is arc length and ϕ is the angle from a chosen fixed direction to the tangent drawn in the

Parasitology

... which deals with those organ-

[illegible]

parasitism is dependency, the loss of freedom to live an independent existence; all degrees of dependency obtain, from transitory symbiosis to complete helplessness. There is a direct correlation between the degree of dependency and the extent of parasitism. Plants are primarily free-living and consequently parasitism is far more prevalent in the animal kingdom. Here it is both extensive and intensive, since few, if any, species are free from attack and since the parasites are often present in enormous numbers. Parasitism has appeared in every phylum of animals, where the individuals themselves are parasites or serve as hosts to parasitic forms. Some parasites are temporary or facultative, but most are obligatory, that is, unable to survive apart from their hosts. Parasitism involves a gradual and progressive adaptation on the part of the parasite, and recovery of an independent status becomes increasingly difficult. It is clear that parasitic species have been derived from free-living ancestors and this mode of life is probably as old and as universal as animal associations themselves.

Practically all animal parasites are invertebrates, and although other groups have parasitic members, most parasitic species are found among the protozoans, the worms, and the arthropods. Certain of these groups, the Sporozoa, Trematoda, Cestoda, and Acanthocephala, are exclusively parasitic. The Sporozoa have no genetic entity and comprise a heterogene...

[illegible]

...quatic vertebrates, and the Digenea, in which the sexual generation in a vertebrate host alternates with two or more asexual generations in an invertebrate host, usually a mollusk. Different species of sexually mature worms live in the intestine, bile ducts and gallbladder, lungs, and blood vessels of their hosts, while the asexual generations live in the hemocoele or blood sinuses of snails and clams. The Cestoda, or tapeworms, are parasites in the digestive system of vertebrates, with larval stages which occur typically in arthropods before vertebrate infestation. A second intermediate host is present in species that infest crustaceans, but in the cyclophyllidean or taenioid tapeworms, only two hosts are involved and a vertebrate may serve as the intermediate host. The Acanthocephala are also parasites of the digestive tract of vertebrates, with larval stages in arthropods, but they are dioecious, do not form strobilas, and are characterized by a protrusible proboscis which is armed with hooks or spines. Supplementing the groups listed, the Nematoda include an enormous number of parasitic species and these worms are of great economic and medical importance.

As organisms become dependent, they undergo significant changes in habits, functions, and physiological requirements. These changes are accompanied by a gradual and progressive reduction of those structures which function most actively in a free-living existence. One after another, organs suffer regressive changes. With enfeeblement and atrophy of sensory and muscular elements, there is a corresponding degradation of the central nervous system. Reduction and loss of the respiratory and circulatory organs are followed by loss of the alimentary tract, and in cestodes, acanthocephalans, and such copepods as *Monstrilla* and *Sacculina thompsonia*, there are no traces of the digestive system. As the other organ systems atrophy and their activity declines, the abundant nourishment and active metabolism of the parasite finds expression in enhanced sexual activity and the production of enormous numbers of eggs and larvae. Reproductive capacity is further augmented by the acquisition of new and accessory methods of reproduction: sporulation, paedogenesis, parthenogenesis, and polyembryony (*see* REPRODUCTION, ANIMAL). The enormous fertility of endoparasitic species has enabled them to overcome the hazards imposed both by their complicated life cycles and by the transfers required by alternation of hosts.

Host animals are more or less seriously affected by parasitic infestations, but if well nourished, they may not manifest patent illness. If undernourished, the host becomes increasingly debilitated and the development or activity of the sexual organs is inhibited; parasitic castration is the ultimate result. Parasites may occlude ducts, produce lesions that permit bacterial invasion, or produce substances toxic to the host. Their presence may be detected by serological tests and by increase in number of eosinophil leukocytes. Examination of blood and feces, however, is the surest method of diagnosing infestation. In most instances, defense mechanisms, humoral or others, are developed against infestation, or at least against superinfection. In certain instances, complete, although temporary, immunity is acquired. See separate articles on the various animal groups; *see also* PARASITIC CASTRATION.

[H.W.S.]

Bibliography: T. W. M. Cameron, *Parasites and Parasitism*, 1956; M. Caullery, *Parasitism and Symbiosis*, 1952; G. LaPage, *Parasitic Animals*, 1951.

Parasitology, medical

That branch of medical microbiology which deals with the relationships between man and those animals which live in or on him. In this definition, as currently used, the terms parasite and animal are not taken precisely for two reasons: (1) many harmless organisms are considered legitimate subject matter of medical parasitology; (2) it is well realized that the distinction between plant and animal may be untenable at the unicellular level

unable to correct such experiences and convictions and test them in the light of reality. The disorder is chronic, but does not lead to a profound or general deterioration. It usually occurs in the mature age group.

The etiology is obscure, although many psychiatrists are impressed by the importance of the same psychological factors which play a role in the etiology of paranoid schizophrenia.

Organic therapy, such as electric convulsive therapy, insulin coma treatment, and psychosurgery, has been of little help. Psychotherapy is very difficult, although once confidence between therapist and patient has been established, some patients have benefited from supportive and analytic psychotherapy. Many patients with paranoid states need to be hospitalized in psychiatric institutions. *See* PSYCHOSURGERY; PSYCHOTHERAPY. [F.C.R.]

Bibliography: J. R. Ewalt, E. A. Strecker, and F. G. Ebaugh, *Practical Clinical Psychiatry*, 8th ed., 1957.

Parapertussis

A disease which resembles whooping cough but which is milder and usually nonfatal. It is caused by a bacterium, *Bordetella parapertussis*, which closely resembles *B. pertussis*, but is distinguished by its utilization of citrate and the production of a brown pigment. Vaccination against pertussis does not protect against parapertussis even though the organisms share several antigens *See* HEMOPHILIC BACTERIA, WHOOPING COUGH. [W F.V.]

Parasitic castration

Destruction of the reproductive organs by parasites. Direct castration occurs when parasites penetrate the sex organs and feed on them, as trematodes do in the gonads of mollusks Indirect castration results when parasites cause the gonads to atrophy Some parasitic barnacles cause the regression of gonads of crabs but do not actually invade the gonads. Changes in the secondary sex characters commonly result

In animals, Crustacea furnish good examples of parasitic castration The gonads may completely atrophy in crabs parasitized by Rhizocephala. More often, the development of the sex gland is retarded. A characteristic effect is the failure of egg cells to deposit yolk. The most noticeable external change in parasitized male crabs is broadening of the abdomen so that it resembles that of a normal female In parasitized females an adult type of abdomen appears precociously but growth of the pleopods is inhibited.

In plants, the term parasitic castration has been employed to describe the action of smut fungi in the grain of wheat, and the conversion of stamens and carpels into petals in plants infested by nematodes.

Some hypotheses attribute the morphological effects of parasitic castration to metabolic disturbances. Others emphasize hormonal relationships or interference with the action of sex genes. *See* RHIZOCEPHALA; SMUT (MICROBIOLOGY). [P.C.R.]

Parasitic oscillation

Undesired oscillations, which may occur in any type of circuit such as an audio-, video-, or radio-frequency amplifier, oscillator, modulator, or pulse waveform generating circuit. For example, it often happens that, with no apparent input signal to an amplifier, an output voltage of considerable magnitude is obtained. The amplifier may be oscillating because some part of the output is inadvertently being fed back into the input. This feedback may result from the output impedance of the power supply. If feedback does occur through the power supply impedance, the oscillations can usually be stopped by the use of appropriately placed decoupling networks. Such a filter is obtained by placing a resistor in series with the plate load and bypassing the resistor to ground with a large capacitor. *See* AMPLIFIER; FEEDBACK CIRCUIT; OSCILLATOR.

Feedback may also occur through the interelectrode capacitance from grid to plate of a tube, through lead inductances, stray wiring, and other paths, which are often difficult to determine exactly Parasitic oscillations are particularly prevalent in circuits where physically large tubes are used, in circuits where tubes or transistors are operated in parallel or push-pull, and in power stages. The frequency of oscillation may be in the audio range but is usually much higher Often it is so high (hundreds of megacycles) that its presence cannot be detected with an oscilloscope. A low-wattage neon bulb insulated from ground may be used as an indicator. When the lamp is brought near that portion of the circuit which is oscillating it will glow

Parasitic oscillations represent a waste of power, a distortion of the desired waveform, or a complete malfunctioning of the circuit Hence, these oscillations must be eliminated This can usually be accomplished by a change in circuit parameters, a rearrangement of wiring, some additional bypassing or shielding, a change of tube or transistor, the use of an rf inductor in the plate circuit, the use of rf chokes in series with filament lead, and other methods. A small resistance (50-1000 ohms) placed in series with a grid and as close to the grid terminal as possible is often effective in reducing high frequency oscillations.

Even if the circuit is not oscillating, an output voltage may be present in a vacuum-tube amplifier in the form of hum from the use of ac heated filaments Some hum may also appear from pickup resulting from stray magnetic fields of the power transformer or from the fields produced by the heater current in the connecting leads. It is also possible to pick up rf signals radiated through space Spurious output voltages caused by vibrations of the electrodes because of mechanical or acoustical jarring of the tube are called microphonics The undesired waveforms discussed in this paragraph should not be confused with true parasitic oscillations. *See* NOISE, ELECTRICAL; SHIELDING, ELECTRICAL. [J.M.L.]

1. In the United States, an estimated 10% of population is infected with *Entamoeba histo-a*, 20% or more of adults have become in-ed with the trichina worm (*Trichinella*) at one or another, a larger number are believed to contracted a *Toxoplasma* infection sometime heir past, perhaps some 7,000,000–8,000,000 ra in the United States carry *Trichomonas nalis*, and so on

general. control measures are available for of the parasitic infections, application of measures depending on extrascientific factors. ng the major scientific problems which remain (1) the mechanism or mechanisms of disease ation by parasites, and (2) the origin of sitism in the different major groups and the ure of the adaptive processes employed in nging over from a free living organism to one suited to parasitic existence in animals.

For diseases due to Protozoa see AMEBIASIS, LANTIDIASIS, CHAGAS' DISEASE, LEISHMANIASIS; LARIA; PNEUMOCYSTOSIS; TRICHOMONIASIS

For diseases due to helminths see ASCARIASIS; ONORCHIASIS, FASCIOLIASIS; FASCIOLOPSIASIS; LARIASIS, GUINEA WORM INFECTION; HETERO-PHYIASIS, HOOKWORM DISEASE, PARAGONIMIASIS, WORM INFECTION; SCHISTOSOMIASIS; STRONGY-LOIDIASIS; TAPEWORM DISEASE, TRICHINOSIS.

[D.W.]

Bibliography: F. M. Burnet, *Natural History of Infectious Disease*, 1953; H. A. Christian (ed.), *The Oxford Loose-leaf Medicine*, vols. 4 and 5, 1920–1942; E. C. Faust and P. F. Russell, *Craig and Faust's Clinical Parasitology*, 1957; T. Smith, *Parasitism and Disease*, 1934

Parasympathetic nervous system

A portion of the autonomic system. In general, its action is in opposition to that of the sympathetic nervous system, which is the other part of the autonomic system. It cannot be said that one system, the sympathetic, always has an excitatory role and the other, the parasympathetic, an inhibitory role, the situation depends on the organ in question. However, it may be said that the sympathetic system, by altering the level at which various organs function, enables the body to rise to emergency demands encountered in flight, combat, pursuit, and pain. The parasympathetic system appears to be in control during such pleasant periods as digestion and rest. The alkaloids pilocarpine and atropine affect parasympathetic activity, pilocarpine excites and atropine inhibits the activity.

Because the parasympathetic transmission of efferent impulses is restricted to cranial nerves III (oculomotor), VII (facial), IX (glossopharyngeal), and X (vagus) and to sacral segments 1, 2, 3, and 4, it is also known as the craniosacral system.

Results of stimulation of the parasympathetic components of the following nerves are listed here.

1. Oculomotor (III) cranial nerve: Eye accommodation in near vision, constriction of the pupil.

2. Facial (VII) cranial nerve: Secretion, vasodilation, and constriction of ducts in the submaxillary and sublingual salivary glands and in the lacrimal, nasal, and buccal glands.

3. Glossopharyngeal (IX) cranial nerve: Secretion, vasodilation, and constriction of ducts in parotid glands.

4. Vagus (X) cranial nerve: Decrease in heart rate, coronary constriction, tracheobronchial constriction, increased smooth-muscle mobility in esophagus, gastrointestinal secretion, internal and external pancreas reduction, increase in gastrointestinal motility, inhibition of ileocolic sphincter, and constriction of pancreatic and biliary ducts.

5. Sacral: Evacuation of rectum and bladder, pelvic vasodilation, erection, and secretion of genital tract glands (prostate and Cowper's). See AUTONOMIC NERVOUS SYSTEM, CRANIAL NERVE, SPINAL CORD; SYMPATHETIC NERVOUS SYSTEM. [E.G.ST.]

Parathyroid gland

An endocrine organ usually associated with the thyroid gland in the neck region and possessed by all vertebrates except the fishes. Embryologically, the glandular primordia are formed in the endoderm of the several pharyngeal pouches and are associated with the similarly derived primordia of the thymus. There may be from one to three pairs of the small glands present in individuals of the various vertebrate classes, although two pairs appear most frequently. They are characteristically within, on, or near the thyroid gland. In response to lowered serum calcium concentration, a hormone is produced which promotes bone destruction and inhibits the phosphorus-conserving activity of the kidneys. The hormone, a simple protein, can be extracted from the whole gland with acidic aqueous solutions. Injections of the extract, called parathormone, cause the calcium concentrations in the blood serum and urine to rise, whereas phosphate concentrations rise in the urine and fall in the serum. Elimination of the glands causes a rapid fall in serum calcium ion concentration and results in a condition of muscular tetany that may be fatal. Intravenous administration of calcium salts or parathormone arrests the tetany.

Embryology. Developmentally, the parathyroids of all vertebrates originate in the endodermal lining of certain pharyngeal pouches. However, differences in details of the site of origin occur among the several vertebrate classes and are summarized in Table 1.

The broader aspects of parathyroid development in the various classes are comparable and differ only in detail, as an example, that which occurs in man is given as representative. At 35–37 days of de[illegible]

of cells that constitute the glandular primordia. One primordium is established on the dorsolateral surface of each of the paired third and fourth pharyngeal pouches. Unlike the primordia of the [illegible]

Phyla involved. If only the most important human parasites are considered, they fall into four different phyla, whereas medical bacteriology and medical mycology, for example, are each concerned with a single phylum only. These phyla are the Protozoa, the Nematoda or roundworms, the *Platyhelminthes* or flatworms, and the Arthropoda, including the insects, ticks and mites. The Spirochaetaceae, frequently included, are treated also as a part of bacteriology. *See* ACARINA; ANOPLURA; BACTERIOLOGY, MEDICAL; CESTODA; DIPTERA; MALLOPHAGA; MYCOLOGY, MEDICAL; NEMATODA; SIPHONAPTERA; SPIROCHAETALES.

Structural and biological variation. The size range varies from a length of 3 μ for Protozoa to 25 m for flatworms, a variation of from 1 to 8,000,000. Structural and biological variation is equally great. There are organisms which are complete in a single cell, multicellular and complex insects with metamorphoses resulting in very different-appearing animals during the stages from egg to adult, and hermaphroditic flatworms which, also multicellular and also developing through dissimilar-appearing larval stages, can direct their development either forward to maturity or backward toward juvenility, and can reproduce at these various stages.

Again, the different organisms show extreme diversity in their localization within the body. Some are exclusively intracellular, most are extracellular only, but some are both. Very nearly every organ, tissue, and fluid of the body is invaded: trypanosomes live in the blood, lymph nodes and cerebrospinal fluid; filaria in the lymph channels and eye fluids, other worms inhabit the intestine, bronchi and bile ducts, and a variety of organisms occur in the heart, brain, liver, spleen, and glands of internal secretion. In these sites, the effect on the tissue varies from complete destruction to almost complete innocuity, and the same parasite at different times may have one or the other of these effects. Finally, the relationships between microorganisms and man include those which are often harmful and those which are usually harmless, and the relationship may be permanent or temporary, the microorganism being known, accordingly, as either an obligatory or a facultative parasite.

Epidemiology. The biological, medical, chemotherapeutic, and control problems in this field are extremely varied, and meaningful generalities are difficult to make. However, all organisms dependent on man face certain common problems, and this provides the basis for unifying concepts. All parasites considered here are human parasites, consequently all must find access to man. Some live on the skin and are passed from person to person, as in scabies. Others penetrate the skin actively, either from the soil, as in hookworm infestation, or from water, as in schistosomiasis. Still others are taken in with ingesta, either as free contaminants of food and drink, such as amebas and many roundworms, or occur within the tissues of common meat foods, as in trichinosis and tapeworm infections. A great variety are inoculated more or less passively by Arthropoda, the latter in such cases being termed the vector of the organism transmitted.

Once it has arrived in the body, the organism must avoid initial destruction, migrate to its preferred site, and then later escape destruction from the specifically directed antibodies and phagocytic mechanisms. It must multiply. Then the final problem of egress must be solved. Some leave the body via the intestinal tract, others utilize the genital secretions or the sputum. In general, the large group which employs Arthropoda as a means of access to the body uses the same method of egress, although some at the end of their cycle in the body become free-living. The problem of egress has only a limited number of solutions; accordingly, almost identical mechanisms have been adopted by microorganisms which are in every other respect entirely different. The process is strikingly analogous to convergence in free-living animals.

Achievements. When compared with the other branches of microbiology, outstanding achievements of this field lie in the field of chemotherapy and in the detailed study of the development of microorganisms in the human body. The long-term prophylaxis of African trypanosomiasis by means of synthetic chemotherapeutic agents, first achieved in the 1920s, is an accomplishment probably still unmatched in medical microbiology. Study of the incubation period, particularly of protozoan diseases, has made it clear that the period immediately following inoculation may not be devoted to simple multiplication of the injected organism, as commonly supposed. Quite the contrary; the inoculated organism is replaced by a very different type of organism, at times by a succession of different kinds of organism. In each of these successive stages, the microorganism not only is morphologically different, but may, and often does, have a different metabolism, different elective sites within the body, and different chemotherapeutic susceptibilities. Failure to grasp the true course of such events leads to erroneous conclusions and fortuitous experimentation.

Impact on human affairs. The effect of parasitological diseases on human history has been enormous. They have exerted an influence in tropical Africa, an area of some 4,500,000 square miles, which has decided the outcome of wars, has been decisive in determining which are the uninhabitable areas, and has kept agriculture and transportation from rising from the man-power to the horse-power level. These consequences are justifiably attributed to the trypanosomiases, human and animal. Other diseases may have had a similar importance, malaria, for example, is believed to have contributed significantly to the factors determining the fall of brilliant civilizations on several continents.

The present importance of the parasitic diseases is still very great. In 1957, it was estimated that there were in the world 200,000,000 malaria cases with 2,000,000 deaths. In 1947, a census of human parasitic worms gave a figure of 2.2×10^9, approximately equal to the human population of the

earth. In the United States, an estimated 10% of the population is infected with *Entamoeba histolytica*, 20% or more of adults have become infected with the trichina worm (*Trichinella*) at one time or another; a larger number are believed to have contracted a *Toxoplasma* infection sometime in their past; perhaps some 7,000,000–8,000,000 women in the United States carry *Trichomonas vaginalis*, and so on.

In general, control measures are available for most of the parasitic infections, application of these measures depending on extrascientific factors. Among the major scientific problems which remain are (1) the mechanism or mechanisms of disease causation by parasites, and (2) the origin of parasitism in the different major groups and the nature of the adaptive processes employed in changing over from a free-living organism to one well suited to parasitic existence in animals.

For diseases due to Protozoa *see* AMEBIASIS; BALANTIDIASIS; CHAGAS' DISEASE; LEISHMANIASIS; MALARIA; PNEUMOCYSTOSIS; TRICHOMONIASIS.

For diseases due to helminths *see* ASCARIASIS; CLONORCHIASIS; FASCIOLIASIS, FASCIOLOPSIASIS; FILARIASIS; GUINEA WORM INFECTION; HETEROPHYIASIS, HOOKWORM DISEASE, PARAGONIMIASIS; PINWORM INFECTION; SCHISTOSOMIASIS; STRONGYLOIDIASIS, TAPEWORM DISEASE; TRICHINOSIS

[D W]

Bibliography: F. M. Burnet, *Natural History of Infectious Disease*, 1953; H A Christian (ed.), *The Oxford Loose leaf Medicine*, vols. 4 and 5, 1920-1942; E. C. Faust and P. F. Russell, *Craig and Faust's Clinical Parasitology*, 1957; T Smith, *Parasitism and Disease*, 1934

Parasympathetic nervous system

A portion of the autonomic system. In general, its action is in opposition to that of the sympathetic nervous system, which is the other part of the autonomic system. It cannot be said that one system, the sympathetic, always has an excitatory role and the other, the parasympathetic, an inhibitory role; the situation depends on the organ in question. However, it may be said that the sympathetic system, by altering the level at which various organs function, enables the body to rise to emergency demands encountered in flight, combat, pursuit, and pain. The parasympathetic system appears to be in control during such pleasant periods as digestion and rest. The alkaloids pilocarpine and atropine affect parasympathetic activity; pilocarpine excites and atropine inhibits the activity.

Because the parasympathetic transmission of efferent impulses is restricted to cranial nerves III (oculomotor), VII (facial), IX (glossopharyngeal), and X (vagus), and to sacral segments 1, 2, 3, and 4, it is also known as the craniosacral system.

Results of stimulation of the parasympathetic components of the following nerves are listed here.

1. Oculomotor (III) cranial nerve. Eye accommodation in near vision, constriction of the pupil.

2. Facial (VII) cranial nerve: Secretion, vasodilation, and constriction of ducts in the submaxillary and sublingual salivary glands and in the lacrimal, nasal, and buccal glands.

3. Glossopharyngeal (IX) cranial nerve: Secretion, vasodilation, and constriction of ducts in parotid glands.

4. Vagus (X) cranial nerve: Decrease in heart rate, coronary constriction, tracheobronchial constriction, increased smooth-muscle mobility in esophagus, gastrointestinal secretion, internal and external pancreas reduction, increase in gastrointestinal motility, inhibition of ileocolic sphincter, and constriction of pancreatic and biliary ducts.

5. Sacral: Evacuation of rectum and bladder, pelvic vasodilation, erection, and secretion of genital tract glands (prostate and Cowper's). *See* AUTONOMIC NERVOUS SYSTEM; CRANIAL NERVE; SPINAL CORD, SYMPATHETIC NERVOUS SYSTEM [E G ST]

Parathyroid gland

An endocrine organ usually associated with the thyroid gland in the neck region and possessed by all vertebrates except the fishes. Embryologically, the glandular primordia are formed in the endoderm of the several pharyngeal pouches and are associated with the similarly derived primordia of the thymus. There may be from one to three pairs of the small glands present in individuals of the various vertebrate classes, although two pairs appear most frequently. They are characteristically within, on, or near the thyroid gland. In response to lowered serum calcium concentration, a hormone is produced which promotes bone destruction and inhibits the phosphorus-conserving activity of the kidneys. The hormone, a simple protein, can be extracted from the whole gland with acidic aqueous solutions. Injections of the extract, called parathormone, cause the calcium concentrations in the blood serum and urine to rise, whereas phosphate

travenous administration of calcium salts or parathormone arrests the tetany.

Embryology. Developmentally, the parathyroids of all vertebrates originate in the endodermal lining of certain pharyngeal pouches. However, differences in details of the site of origin occur among the several vertebrate classes and are summarized in Table 1.

The broader aspects of parathyroid development in the various classes are comparable and differ only in detail; as an example, that which occurs in man is given as representative. At 35–37 days of development, the endodermal cells that are destined to become

become

of cells

One pri

surface

fourth pharyngeal pouches. Unlike the primordia of the fourth pouches, those of the third are intimately associated with thymic primordia (thymus III). Para-

migrates caudally. When the inferior border of the developing thyroid is reached, the parathyroid primordium is released, while thymus III continues its migration. In consequence of this movement, parathyroid III comes to lie inferior to parathyroid IV. Occasionally fragmentation of parathyroid III occurs during migration, the fragments later developing into isolated masses of functional tissue, accessory parathyroids. For this reason it is often difficult or may even be impossible to render an animal entirely free of parathyroid tissue by surgical means.

Table 1. Comparative embryology of parathyroid glands

Vertebrate class, order, and examples	Location of parathyroid primordia: Pharyngeal pouch*	Location of parathyroid primordia: Portion of pouch
Pisces		
Cyclostomes		
Lamprey	?	
Chondrichthyes		
Sharks, skates, and rays	Not known to occur	
Osteichthyes		
Perch, trout	Not known to occur	
Amphibia		
Urodeles		
Salamanders	II(1), III, IV	Ventral
Anura		
Frogs, toads	II(3), III, IV	Ventral
Reptilia		
Sauriens		
Lizards	II†, III, IV†	Ventral
Serpentes		
Snakes	II(?), III, IV, V†	Ventral
Chelonia		
Turtles	III, IV	Ventral
Loricata		
Crocodiles	III	Ventral
Aves		
Galliformes		
Fowl (chicken)	III, IV, V(2)	Ventral
Others	III, IV	Ventral
Mammalia		
Monotremes		
Platypus	II, III, IV	Ventral
Marsupialia		
Opossum	III, IV	Dorsal
Insectivores		
Moles	III, IV(1)	Dorsal
Chiroptera		
Bats	III, IV	Dorsal
Primates		
Monkeys, apes, man	III, IV	Dorsal
Rodentia		
Lagomorpha (rabbits)	III, IV	Dorsal
Muridae (rats, mice)	III, IV†	Dorsal
Caviidae (guinea pig)	III, IV	Dorsal
Carnivores		
Canidae (dogs)	III, IV	Dorsal
Felidae (cats)	III, IV	Dorsal
Perissodactyla (horse)	III, IV	Dorsal
Artiodactyla		
Suidae (swine)	III, IV†	Dorsal
Bovidae		
Sheep, goats	III, IV	Dorsal
Cattle	III, IV	Dorsal

* Numbers in parentheses indicate frequency of occurrence: 1, frequent; 2, occasional; 3, seldom

† Parathyroid primordium is either not formed or undergoes involution shortly after formation

During the fourth to fifth months, chief cells appear and a sinusoidal circulatory pattern is established among the interconnecting cords of cells. The onset of functional activity probably coincides with the appearance of this histological pattern. Variation in size, number, and position of matured glands are known to occur, but complete absence is quite rare and is usually associated with other abnormalities of such severity as to ensure fetal or neonatal death.

Anatomy. Parathyroid glands occur in all vertebrates with the exception of the fishes, although cells that appear to be homologs of parathyroid cells are found in cyclostomes at the dorsal and ventral ends of all pharyngeal pouches. Whether they function as parathyroid tissue is, however, unknown. Data relating to the details of the comparative anatomy of these glands are summarized in Table 2.

In man, there are typically four glands situated as shown in Fig. 1*a*; however, the typical number varies between three and six, with four appearing about 80% of the time, whereas the location varies to the extent shown in Fig. 1*b*. The extent of variation in primates generally is greater than that observed in all other mammals.

Macroscopically, the glands (or more properly, glandules) appear as oval, round, or irregularly shaped objects, generally having a smooth, highly vascularized surface, although the latter may be slightly roughened (horse) or finely serrated (cattle, swine). Both size and weight are quite variable and appear to have no apparent correlation with total body weight. The color varies from yellowish-pink through brown-red to greyish-red or white, *depending upon the proportion of fat cells to secreting cells*, and the amount of lipid contained in the latter.

Histology. The histological structure of the parathyroids is based upon nonsecretory supportive or stromal elements, and secreting or parenchymal elements.

The glandules are enclosed in a smooth connective tissue capsule from which trabeculae, bearing lymphatics, blood vessels, and nerves, emerge to penetrate the gland and to divide its substance into a roughly lobular pattern. The blood vessels are connected to a rich capillary network, characteristic of the endocrine glands generally, that is entwined in reticular fibers found among the parenchymal cells. In addition to these elements, fat cells also occur and increase in number with advancing age.

Parenchymal cells are arranged in interconnecting cords and clumps of cells as well as in occasional follicular masses enclosing a gelatinous pro-

naterial called colloid; the latter resembles id colloid in appearance, but does not contain and appears to have no comparable physio- l activity

There are two general types of parenchymal cell, the chief cell (also called the principal or chromophobe cell) and the oxyphil cell (eosinophil or acidophil). These cells are classified according to

2. Comparative anatomy of parathyroid glands

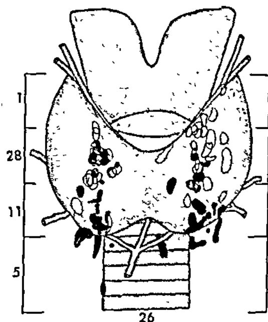
rate class, r, and examples	Number of glandules, location, and relations*
bia	
dèles	
alamanders	2-3 pairs of glandules separate from thyroid, ventral to thymus, lateral to arterial arches†
ura	
Frogs, toads	2-3 pairs of glandules lie laterally next to EJ; gill remnant is immediately dorsal in frog, just ventral in toad†
ilia	
urians	
Lizards	One pair of glandules, III only; usually attached to thymus
rpentes	
Snakes	2 pairs of glandules which may be II and III or III and IV (probably latter), if so, III is near temporomaxillary joint next to bifurcation of CC, IV lies between thymus III and IV
Chelonia	
Turtles	2 pairs of glandules, III at cranial end, IV at caudal end of thymus
Loricata	
Crocodiles	One pair of glandules, III *only, next* to thymus, near origin of collateral cervical artery
es	
Galliformes	
Fowl (chicken)	2 pairs of glandules, III at inferior pole of thyroid, IV *next* to III, when fused, connection may be more or less tenuous†
Others	In birds other than fowl, there is wide variation in number and position, *it* is usually associated with thyroid or CC in some way
Mammalia	
Monotremes	
Platypus	3 pairs of glandules, II at bifurcation of CC, *III* at dorsal end of thymus, IV *on lateral* surface of thyroid
Marsupialia	
Opossum	2 pairs of glandules, III medial of branching of CC, IV connected with or embedded in thymus IV, also found between the thyroid and CC
Insectivores	
Moles	One pair of glandules, III near middle CC
Chiroptera	
Bats	2 pairs of glandules, III (larger) on dorsal outer surface of thyroid, IV (smaller) on inner (tracheal) surface of thyroid

Vertebrate class, order, and examples	Number of glandules, location, and relations*
Primates	
Monkeys, apes, man	2 pairs of glandules; *in monkeys and apes*, extent of variation in number and position exceeds that observed in other mammals, generally, an external pair (III) related to thymus, an internal pair buried in the thyroid
Lagomorpha (rabbits)	2 pairs of glandules, III associated with thyroid, either at superior or inferior pole of lateral lobes, free or sunken into a depression on thyroid, often more closely related to CC, IV at middle of lateral lobe of thyroid, exceptionally at inferior pole
Muridae (rats, mice)	One pair of glandules, III only, usually in cavity on lateral surface of thyroid, cavity usually more superficial in mouse than rat
Caviidae (guinea pig)	2 pairs of glandules, III usually removed from thyroid, sometimes vestigial, IV is embedded in thyroid
Carnivores	
Canidae (dogs)	2 pairs of glandules, III on or in front of lateral cranial surface of thyroid, sometimes more or less embedded in surface, IV is small, deeply buried in thyroid, position variable
Felidae (cats)	2 pairs of glandules, III similar to dog, IV lies near middle of thyroid nearer to tracheal than outer surface
Perissodactyla (horse)	2 pairs of glandules, III near (within 1 cm), on or partially embedded in upper medial border of thyroid, occasionally on tracheal surface, IV is entirely embedded in thyroid
Artiodactyla	
Suidae (swine)	One pair of glandules, III only, usually well removed from thyroid, either in thymus or with terminal branches of CC
Bovidae	
Sheep, goats	2 pairs of glandules, III near bifurcation of CC, IV may be either free on surface of thyroid or deeply embedded
Cattle	2 pairs of glandules, III near bifurcation of CC, IV usually lies on inner tracheal surface of thyroid

* Positions given are subject to extensive variation. III or IV used to designate glandules according to pharyngeal ... of origin. In mammals, III is external to thyroid, IV to a greater or lesser extent, within its substance. EJ is ...ternal jugular vein, CC is common carotid artery.

† May occur as separate lobes, but *frequently two lobes are fused or joined more or less tenuously.*


(a)



(b)

Fig 1 (a) Common positions of the human parathyroid glands on the posterior aspect of the thyroid (*from W H. Hollinshead, Anatomy of the endocrine glands, The Surg. Clin. North Am., 32(4):1115–1140, 1952*). (b) Composite diagram, showing the thyroid as a transparent structure, that demonstrates extent of positional variation of the parathyroids in 25 cadavers. Brackets and figures indicate total number of glandules found in each region. Parathyroid IV shown in outline, parathyroid III in solid black, glandules which could be either are shaded (*from W. F. Heinbach, Jr., A study of the number and location of the parathyroid glands in man, Anat Record, 57(3):251–261, 1933*).

the affinity of their cytoplasm for acidic or basic biological stains.

The chief cells, which represent the larger part of the parenchymal cell population, are small and possess a clear cytoplasm that is weakly to moderately acidophilic A number of variants of the cell occur and it has been suggested that they represent phases in a cycle of secretory activity that is concerned with the production of the parathyroid hormone.

Oxyphil cells, however, make up only a small part of the parenchyma and have no apparent function. They may, indeed, be only older, postsecretory stages of the chief cell. They are appreciably larger than chief cells and have a moderately to strongly acidophilic cytoplasm and small, dark-staining nuclei Oxyphils are seldom found as single cells but occur most frequently in clusters.

Chief cells appear in parathyroids of all vertebrates Oxyphil cells, however, appear in the human only after the fourth to seventh year, and in the older macaque monkey and cattle. They have not been described in other forms. These observations support the thesis that the chief cell is the true functional cell of the organ [W E.D.]

PHYSIOLOGY

The parathyroids were formerly regarded as accessory thyroid glands, and in early studies on thyroidectomy of cats and dogs, the parathyroids were also removed; the resulting nervous irritability, tetany, convulsions, and ultimate death were erroneously attributed to thyroid deficiency It is now known that the neuromuscular symptoms are not due to thyroid deficiency but to the absence of a parathyroid hormone This hormone has not been obtained in homogeneous form, but it appears to be a protein. It is effective only when given parenterally.

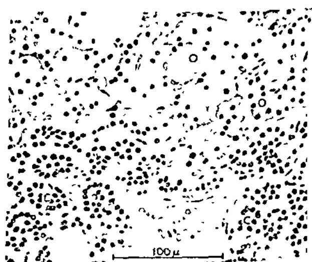


Fig 2. Histology of the normal human parathyroid gland. (*From J R. Gilmour, The normal histology of the parathyroid glands, J Pathol and Bacteriol, 48(1):187–222, 1939*)

Hypoparathyroidism. This condition is sometimes observed clinically and may be produced experimentally by surgical ablation of the glands. The severity of the symptoms varies widely with the species, age, reproductive status, and the nature of the diet.

The most prominent symptom of parathyroid deficiency is hyperirritability of the nervous system which accounts for the latent or manifest tetany. As the diffusible calcium in the plasma falls, the operated animals typically develop generalized convulsions, the tonic spasms eventually spreading to all the muscles. Death frequently results from spasm of the respiratory muscles. The symptoms may be relieved by the administration of Ca^{++} salts. In some species, particularly herbivores, parathyroidectomy produces only mild symptoms and convulsive seizures may be absent unless the biochemical status is aggravated by pregnancy, fasting, low calcium diets, or alkalosis.

The main biochemical changes in parathyroid deficiency are low serum calcium levels, high serum phosphate, and reduced urinary elimination of both calcium and phosphate. The serum calcium normally ranges from 9 to 11 mg%. From 50 to 70% of the serum calcium is ultrafiltrable and diffusible; the remainder is bound to protein. After parathyroidectomy, the total calcium falls to 5–7 mg%, and the diffusible calcium ranges from 0 to 2 mg%.

Parathyroid deficiency may be ameliorated by feeding diets high in calcium and low in phosphate, injecting parathyroid hormone, or administering vitamin D and related substances. Although vitamin D promotes the absorption of calcium from the intestine and facilitates the excretion of phosphate by the kidneys, no relationship is known to exist between vitamin D and the hormone. Parathyroid hormone has no effect in relieving vitamin D deficiency.

Hyperparathyroidism. This condition occurs clinically and may be produced by giving parathyroid extract. The biochemical changes are opposite to those observed in parathyroprivic subjects. The hormone induces a prompt increase in the rate of phosphate excretion and the serum phosphate begins to fall. The blood calcium rises more slowly, and after some lapse of time the renal elimination of calcium is increased. The administration of moderate amounts of hormone over long periods leads to a negative balance of calcium and phosphate, both minerals being mobilized from the bones. Consequently, the demineralized skeleton is easily fractured and bent. The high level of calcium in the body fluids leads to formation of kidney stones and metastatic calcium deposits in other soft tissues. The syndrome of hyperparathyroidism in laboratory animals simulates the clinical condition called osteitis fibrosa cystica.

Secretory regulation and action. Regulation of parathyroid secretion seems to operate on the principle of a feedback mechanism. Diminishing levels of serum calcium stimulate the parathyroids to increase their output of hormone which in turn elevates the blood calcium by mobilizing the calcium stores of the skeleton. If there is any nervous control of the parathyroids, it is very meager. No convincing evidence has been forthcoming for a postulated parathyrotropic hormone from the anterior hypophysis.

The site of action of the parathyroid hormone has not been finally settled. Three viewpoints have been proposed, the renal (phosphaturic) theory, the bone theory, and the citrate theory.

Renal theory. The renal theory is based on the finding that hypoparathyroid subjects respond to parathyroid hormone by a prompt urinary excretion of phosphate. By directly acting on the kidney, the hormone promotes the elimination of phosphate, resulting in a lowering of serum phosphate, withdrawal of phosphorus from the skeleton, and mobilization of calcium, thus elevating the levels of serum calcium.

Bone theory. The bone theory proposes that the primary effect of parathyroid hormone is stimulation of osteoclastic activity of bone. In nephrectomized animals, parathyroid extracts continue to produce their characteristic effects—elevation of serum calcium and bone resorption. When parathyroid grafts are placed in close proximity to a bone, they induce a localized resorption of bone substance. The preponderance of evidence supports the view that the parathyroid glands function in calcium homeostasis through a primary effect of the hormone on bone.

Citrate theory. According to this theory, the hormone causes an increase of the concentration of citrate ion in the serum which, in turn, exerts a powerful solubilizing action on hydroxylapatite and on bone minerals. None of these theories has been established definitively. It appears likely that several or all of these mechanisms may be operative simultaneously. See BONE; DIGESTIVE SYSTEM; ENDOCRINE SYSTEM; VITAMIN D [C.D.T.]

Bibliography: L. Bolk et al. (eds.), *Handbuch der vergleichenden Anatomie der Wirbeltiere*, vol 3, 1937; D. H. Copp, Calcium and phosphorus metabolism, *Am. J. Med.*, 22(2):275–285, 1957; G. Pincus and K. V. Thimann (eds.), *The Hormones*, vol 1, 1948, vol 3, 1955; W. H. von Moellendorff, *Handbuch der mikroskopischen Anatomie des Menschen*, vol 3, pt 2, 1939; B. H. Willier, P. Weiss, and V. Hamburger, *Analysis of Development*, 1955.

Paratyphoid fever

An acute, infectious disease of man caused by paratyphoid bacilli. The gram negative, rod shaped bacilli frequently involved are *Salmonella paratyphi A*, *S. paratyphi B*, and *S. typhimurium* and, in some parts of the world, *S. paratyphi C* is also implicated. Those infections caused by *S. cholerae suis* are distinguished by clinical severity and high mortality. The mortality due to infections by this organism is over 20%. In contrast, a mortality in the order of 3–5% is usually found with the infections caused by other *Salmonellae*. Paratyphoid fever is clinically indistinguishable from typhoid fever (see TYPHOID FEVER). The disease is trans-

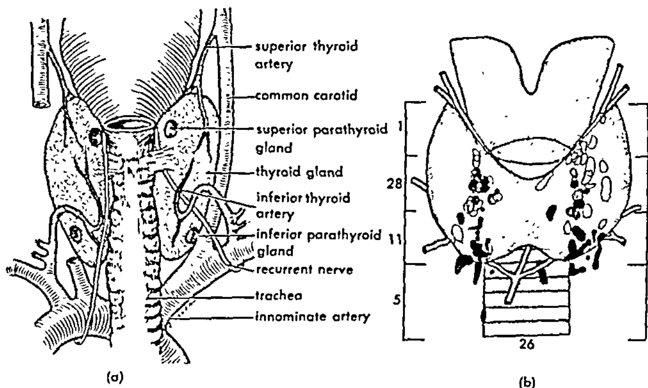


Fig 1 (a) Common positions of the human parathyroid glands on the posterior aspect of the thyroid (from W H Hollinshead, *Anatomy of the endocrine glands, The Surg Clin North Am*, 32(4) 1115–1140, 1952) (b) Composite diagram, showing the thyroid as a transparent structure, that demonstrates extent of positional variation of the parathyroids in 25 cadavers Brackets and figures indicate total number of glandules found in each region. Parathyroid IV shown in outline, parathyroid III in solid black; glandules which could be either are shaded (from W. F. Heinbach, Jr., *A study of the number and location of the parathyroid glands in man, Anat Record*, 57(3):251–261, 1933)

the affinity of their cytoplasm for acidic or basic biological stains.

The chief cells, which represent the larger part of the parenchymal cell population, are small and possess a clear cytoplasm that is weakly to moderately acidophilic A number of variants of the cell occur and it has been suggested that they represent phases in a cycle of secretory activity that is concerned with the production of the parathyroid hormone

Oxyphil cells, however, make up only a small part of the parenchyma and have no apparent function They may, indeed, be only older, postsecretory stages of the chief cell They are appreciably larger than chief cells and have a moderately to strongly acidophilic cytoplasm and small, dark-staining nuclei Oxyphils are seldom found as single cells but occur most frequently in clusters

Chief cells appear in parathyroids of all vertebrates. Oxyphil cells, however, appear in the human only after the fourth to seventh year, and in the older macaque monkey and cattle. They have not been described in other forms These observations support the thesis that the chief cell is the true functional cell of the organ. [W.F.D]

PHYSIOLOGY

The parathyroids were formerly regarded as accessory thyroid glands, and in early studies on thyroidectomy of cats and dogs, the parathyroids were also removed; the resulting nervous irritability, tetany convulsions, and ultimate death were erroneously attributed to thyroid deficiency. It is now known that the neuromuscular symptoms are not due to thyroid deficiency but to the absence of a parathyroid hormone. This hormone has not been obtained in homogeneous form, but it appears to be a protein It is effective only when given parenterally.

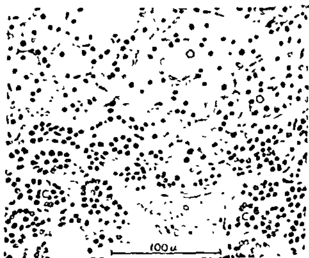


Fig 2 Histology of the normal human parathyroid gland (From J R Gilmour, *The normal histology of the parathyroid glands, J. Pathol and Bacteriol*, 48(1):187–222, 1939)

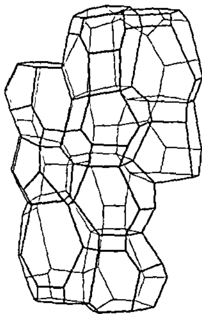


Fig 1 A group of cells from parenchyma of *Asparagus* root, illustrating regular arrangement of cells in a compact tissue (*Photograph by R. L. Hulbary*)

Variations. Typical parenchyma occurs in pith and cortex of roots and stems as a relatively undifferentiated tissue composed of polyhedral cells that may be more or less compactly arranged and show little variation in size or shape (Fig 1). The mesophyll, that is, the tissue located between the upper and lower epidermis of leaves, is a specially differentiated parenchyma called chlorenchyma because its cells contain chlorophyll in distinct chloroplastids (Figs. 2, 3).

This chlorenchymatous tissue is the major locus of photosynthetic activity and consequently is one of the more important variants of parenchyma (*see* PHOTOSYNTHESIS). Specialized secretory parenchyma cells are found lining resin ducts and other secretory structures (*see* SECRETORY STRUCTURES, PLANT).

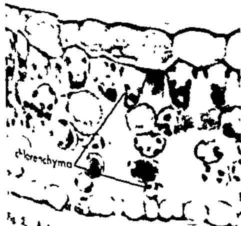


Fig 2 A transverse section of a leaf of *Tolmiea menziesii*, showing mesophyll composed of chlorenchyma with prominent intercellular spaces (*Photograph by R B Wylie*)

Cell contents. Parenchyma cells include the protoplast, common to all living cells, usually with a large vacuole located in the center and the cytoplasm and nucleus next to the cell wall. The kinds of inclusions are partly related to the position of the cell in the plant Thus cells in leaf mesophyll always have chloroplasts. Cells in root cortex commonly contain grains of storage starch Cells in pith and cortex of stems contain starch grains and crystals, usually of calcium oxalate. Substances in crystal form are usually interpreted as waste products and are lost with leaf fall, shedding of bark, or decay of pith Parenchyma in flower parts or fruits may have yellow or red plastids in the cytoplasm or purple or blue anthocyanin pigments in the vacuole (*see* ANTHOCYANIN).

Cell walls. Parenchyma cells are typically thin and contain cellulose, hemicelluloses and pectins (*see* CELLULOSE, HEMICELLULOSE; PECTIN). They may become thick, hard, and lignified as, for example, in wood parenchyma (*see* CELL WALLS IN PLANTS)

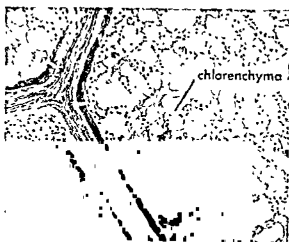


Fig 3 A section of the leaf of *Olearia rami* cut parallel to the surface It shows mesophyll composed of chlorenchyma with prominent intercellular spaces (*Photograph by R B. Wylie*)

Cell shape. The three-dimensional cell shape of parenchyma is frequently a function of cell arrangement In a compact tissue, *parenchyma cells* are almost isodiametric and polyhedral with an average of approximately 14 sides in contact with 14 other cells In more loosely arranged tissues with many air spaces, the average number of contacts with adjoining cells is reduced Parenchyma cells may deviate from the isodiametric shape. Those in the vascular tissues are often conspicuously elongated Parenchyma cells in the ...
and in other
may be vario
CYTOPLASMIC
CORTEX, PLANT, EPIDERMIS, PLANT, FLOWER (BOTANY), FRUIT (BOTANY), LEAF (BOTANY);

mitted in a manner similar to typhoid fever, for example, by contamination of food and water with organisms from active cases or from carriers. For data on the serotypes of *Salmonella*, diagnosis, and epidemiology *see* SALMONELLA. [A.J.W.]

Bibliography: A. J. Weil and I. Saphra, *Salmonellae and Shigellae*, 1953.

Paratyphoid gastroenteritis

The infection by the paratyphoid bacilli, *Salmonella*, restricted to the gastrointestinal tract. Any of the numerous serotypes of *Salmonella* may be the causative agent. However, *S. typhimurium* is the one most frequently found. The infection is characterized by fever, diarrhea, vomiting, abdominal pain, and sometimes prostration and dehydration. The infection is spread by contamination of water and food by organisms from active cases or carriers of *S. typhi murium*. *See* BACTERIOLOGY, MEDICAL; EPIDEMIOLOGY, SALMONELLA.

[A.J.W.]

Parazoa

A subkingdom, in one system of classification, of the animal kingdom in contrast to the subkingdoms Protozoa, Mesozoa, and Metazoa. In another taxo-

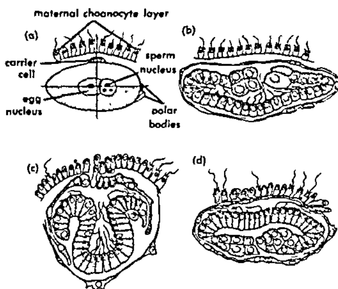


Fig. 1. Development of Sycon. (a) Fertilized egg (b) Stomoblastula. (c) Blastula in process of eversion (d) Inverted blastula. (*After Duboscq and Tuzet, 1937*)

nomic scheme, the animal kingdom is divided into ... Metazoa In this ... nsidered to com- ... 'arazoa, and Eumetazoa. Development of the sponges (Fig. 1) supports the interpretation that the sponges be separated from other Metazoa to form the subkingdom Parazoa. One phylum, the Porifera, is recognized as constituting this group. The similarity in structural details between the choanocytes of sponges and choanoflagellates (Fig. 2) has led many zoologists to consider these protozoans to be ancestral

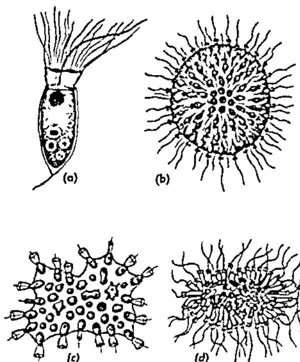


Fig 2. (*a*) *Salpingoeca infusionum*, a choanoflagellate with cytoplasmic filaments arising from the colla (*after Griessmann, 1914*) (*b*) *Sphaeroeca volvox*, a colonial planktonic choanoflagellate (*from Hollande 1952, after Lauterborn, 1894*). (*c*) *Proterospongia haeckeli* (*after Kent, 1881*). (*d*) Free-swimming aggregate of cells of *Grantia compressa* which formed after dissociation of adult cells (*after Tuzet, 1945*).

to the sponges. *See* ANIMAL KINGDOM; EUMETAZOA; MESOZOA; METAZOA; PORIFERA; PROTOZOA.

[C.B.C.]

Parenchyma

A ground tissue chiefly concerned with the manufacture and storage of food The primary functions of plants, such as photosynthesis, assimilation, respiration, storage, secretion, excretion—those associated with living protoplasm—proceed mainly in parenchyma cells (*see* PLANT PHYSIOLOGY). Parenchyma is frequently found as a homogeneous tissue in stems, roots, leaves, and flower parts. Other tissues, such as sclerenchyma, xylem, and phloem, seem to be embedded in a matrix of parenchyma; hence is derived the use of the term ground tissue with regard to parenchyma The parenchyma cell is one of the most frequently occurring cell types in the plant kingdom. *See* PLANT ANATOMY

Origin. Parenchyma cells are differentiated in the primary plant body (shoot, root) from apical growing zones or meristems (*see* MERISTEM, APICAL). The ground meristem that initiates pith, cortex, and leaf mesophyll is the seat of development of parenchyma. In the secondary plant body parenchyma cells are derived from the vascular cambium and cork cambium in the form of ray tissues, xylem and phloem parenchyma, and phelloderm (*see* MERISTEM, LATERAL).

ity (quantum mechanics)

hysical property of a wave function which specifies the wave function's behavior under simultaneous reflection of all spatial coordinates of the e function through the origin, that is, when x eplaced by $-x$, y by $-y$, and z by $-z$. If the wave ction ψ satisfies

$$\psi(x,y,z) = \psi(-x,-y,-z)$$

is said to have even parity. If, on the other hand,

$$\psi(x,y,z) = -\psi(-x,-y,-z)$$

wave function is said to have odd parity. These o expressions can be combined in the form

$$\psi(x,y,z) = P\psi(-x,-y,-z)$$

here $P = \pm 1$ is a quantum number having only e two values $+1$ (designated as even parity) and 1 (odd parity). The physical property defined by is quantized and is called parity. More precisely, arity is defined as the eigenvalue of the operation f space inversion. Parity is a concept that has no eaning in classical particle physics, because it n be defined for a particle only in terms of the chrödinger wave function ψ.

Any eigenstate of a localized object, such as an tom or an atomic nucleus, has a definite parity. orresponding to the fact that the wave function of complex system is the product of the wave funcon of the coordinates of the subsystems into which the system may be subdivided times the internal wave functions of those subsystems, the parity of the system is the product of the parity of the wave function of the coordinates of the subsystems times the intrinsic parities of these subsystems. *See* QUANTUM MECHANICS.

Conservation of parity means that if the wave function describing the *initial* state of a system has even (odd) parity, the wave function describing the final state has *even* (*odd*) parity. Parity conservation is a symmetry law; *see* SYMMETRY LAWS (PHYSICS). It is now well established that parity is not conserved in any of the weak interactions (for example, β decay in radioactivity). In all other interactions parity is conserved, and this fact has an important bearing on atomic transitions and nuclear reactions. In reactions in which particles are created or destroyed, the effect of conservation of parity depends upon the intrinsic parities of the particles. *See* ELEMENTARY PARTICLE.

Since according to quantum theory, a wave function $\psi(x,y,z)$ provides a complete description of a particle (see QUANTUM THEORY, NONRELATIVISTIC), every elementary particle has associated with it an intrinsic parity. Nuclear energy states are characterized by a definite parity (which may be different for different energy states of the same nucleus), and the conservation or nonconservation of parity has an important bearing on nuclear reactions. Operators representing dynamical variables can also be classified in terms of the parity concept, depending upon how they are affected by an inversion of their spatial coordinates.

Parity conservation. The conservation of parity is a consequence of the inversion symmetry of space. To show this formally, let $\mathcal{P}$ be the parity operator which inverts space, that is, $\mathcal{P}$ acting on a wave function yields the wave function at the inverse point of space, $\mathcal{P}\psi(\mathbf{r}) = \psi(-\mathbf{r})$; similarly, for an operator A, $\mathcal{P}A(\mathbf{r})\mathcal{P}^{-1} = A(-\mathbf{r})$. The statement that the world is symmetrical to inversion means that the Hamiltonian H after inversion is the same as before, that is, $\mathcal{P}H\mathcal{P}^{-1} = H$; and thus $\mathcal{P}H - H\mathcal{P} = [\mathcal{P},H] = 0$. Since $\mathcal{P}$ commutes with H, it is a constant of the motion. As for the eigenvalues of $\mathcal{P}$, note that $\mathcal{P}^2 = 1$, from which it follows that the possible eigenvalues of $\mathcal{P}$ are $+1$ or -1. That is, an eigenfunction of $\mathcal{P}$ satisfies $\mathcal{P}\psi_\pm(\mathbf{r}) = \psi_\pm(-\mathbf{r}) = \pm\psi_\pm(\mathbf{r})$, where the upper (lower) sign indicates an eigenfunction of positive (negative) parity, also known as even (odd) parity.

Thus parity would be conserved if the statement of physical laws were independent of the handedness of one's coordinate system. Of course, the fact that most people are right-handed is not a physical law but an accident of evolution; there is nothing in the laws of physics which favors a right-handed to a left-handed man. The same holds for optically active (stereochemical) organic compounds. However, the statement that the neutrino is left-handed is a physical law (see NEUTRINO).

All the strong interactions between elementary particles (for example, nuclear forces) and the electromagnetic interactions are symmetrical to inversion, so that parity is conserved by these interactions. As far as is known, only the β-interactions (which involve neutrinos) and the other weak interactions are not symmetrical to inversion and do not conserve parity. The weak interactions contribute negligibly to all processes except the decays of elementary particles (including β-decay of nuclei), so that in all other processes parity is conserved.

Orbital parity. Since parity is conserved in strong and electromagnetic interactions, it is termed a good quantum number, and an energy eigenstate (unless it is degenerate) must be an eigenstate of parity. The parity of a one-particle state of orbital angular momentum l is given by $P = (-)^l$, that is, even $(+1)$ for $s, d, \ldots$ states and odd (-1) for $p, f, \ldots$ states. Thus the deuteron, whose state is a linear combination of ${}^3S_1$ and ${}^3D_1$, has even parity; there cannot be any admixture of ${}^1P_1$. The orbital parity of an n-particle system is the product of the parities of the $n - 1$ relative orbital angular momentum states: $P_{orb} = (-)^{l_1 + \cdots + l_{n-1}}$. Thus the parity of an atom is the product of the parities of the one-electron orbital wave function; all configurations which mix must have the same parity. The Laporte rule of atomic spectroscopy, which states that an electric dipole transition can occur only between states of opposite parity, depends on the fact that the electric dipole radiation field has odd parity.

PHLOEM; PITH; ROOT (BOTANY); SCLERENCHYMA; STEM (BOTANY); XYLEM. [R.L.HU.]

Bibliography: *See* PLANT ANATOMY.

Paresis, general

An inflammatory and degenerative disease of the brain caused by an infection with *Treponema pallidum*. It is also called syphilitic meningoencephalitis (*see* SYPHILIS). The disease is characterized by psychotic behavior with a profound and progressive intellectual and emotional deterioration and neurological changes. The disease, unless treated, takes a rapid downhill course and results in death. *See* DEATH, PSYCHOSIS

The symptoms consist of a profound impairment of abstract thinking and judgment. Patients become emotionally unstable, crude, and often excited. They develop silly, grandiose and expansive, or, more rarely, paranoid delusions (*see* DELUSION). Most patients are euphoric, a few depressed Their memory and ability to recall are markedly impaired. Neurological signs are changes in the shape and reactions of the pupils such as reaction to accommodation or distance but no reaction to light (Argyll Robertson's sign), disturbances of speech and writing, changes of deep tendon reflexes, changes in facial expression, and, particularly in the malignant forms, trophic (nutritional, digestive, and assimilative) changes. The most common types of personality change are the simple euphoric deterioration, the grandiose expansive types, and combinations with tabes dorsalis. Tabes dorsalis is a form of neurosyphilis in which there is a selective degeneration in the posterior roots of the spinal nerves and the posterior columns of the spinal cord

Diagnosis is usually based on the neuropsychiatric examination and confirmed by serological findings in blood and spinal fluid (*see* SEROLOGY) Differential diagnosis against other organic reactions is simple; the differentiation from other syphilitic diseases of the central nervous system is more difficult. The post mortem pathology of the brain shows characteristic signs of luetic, inflammatory, and degenerative changes *See* INFLAMMATION, PATHOLOGY.

With the control and treatment of syphilis in Western countries, general paresis is not an important disease. In primitive and nontropical countries where malaria does not occur, general paresis is still quite common

Treatment of general paresis was at one time inoculation with tertian or quartian malaria (*see* MALARIA). This treatment, for which J. W. von Jauregg received the Nobel Prize, has been largely replaced by treatment with massive doses of penicillin (*see* PENICILLIN). Prevention consists of prevention of a syphilitic infection or, if an infection has occurred, in early and effective treatment of syphilis. *See* PSYCHOSIS. [F.C.R.]

Bibliography: J. R. Ewalt, E. A. Strecker, and F. G. Ebaugh, *Practical Clinical Psychiatry*, 8th ed., 1957.

Paresthesia

One of several designations given to abnormally intense and disagreeable pain arising from apparently trivial stimulation. Other terms used to identify the same symptom are spontaneous pain, overaction or overresponse, paradoxic pain, hyperpathia, protopathic pain, dysesthesia, and central pain. Indeed, there are more different names available than there are good descriptions of the pain.

Whatever its designation, the circumstance leading to such severe pain is usually sensory nerve injury. Upon recovery from the wound there may be left a skin area having anesthesia at its core bordered by paresthesia in a zone lying between it and normal skin. If the injury is sufficiently circumscribed there may be no anesthesia, only the local intermediate zone.

The ugly pain is not the result of hyperesthesia in the usual sense of that term (*see* PAIN, DEEP). The threshold may even be raised considerably. However, once pain is aroused it is abnormally strong, tends to persist after the removal of the stimulus, radiates into other areas, and has a diffuse localization. Henry Head, a British neurologist, the first of many to perform the experiment of severing a cutaneous nerve and following the regeneration of sensitivity in the affected region, referred to the pain of the intermediate zone as wicked pain.

Recent histological evidence strongly suggests that protopathic pain (paresthesia) always results from a reduction of sensory nerve endings in the affected region Where there is the normal overlap of terminations interdigitating from different fibers, pain from pinprick has a normal, more subdued character. *See* PAIN, CUTANEOUS. [F.A.G.]

Bibliography. E G Boring, *Sensation and Perception in the History of Experimental Psychology*, 1942

Parietales

A large order of the plant subclass Dicotyledoneae. Since there is little or no agreement as to natural relationships in this order, the Parietales are generally regarded as an artificial assemblage, the families being variously distributed by many authors. A major characteristic of the group is parietal placentation (ovules attached to the wall of the ovary). The order includes 23 families having 334 genera and 6960 species The majority of the families are tropical and subtropical shrubs and trees. Here belong a large number of both ornamental and useful plants including tea, camellias, St. John's wort, mangosteen (an important edible fruit of the tropics), tamarisk, ocotillo, rockrose, arnotto (source of a dye used to color foods), hydnocarpus (source of chaulmoogra oil used as a remedy for leprosy), passion flower, papaya (one of the most highly prized fruits of the tropics), violets, and the begonias. *See* TEA; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM. [P.D.S.]

It was at first somewhat disconcerting to find parity not conserved, for that seemed to imply a handedness of space which would then not be the empty thing which (since the demise of the ether hypothesis) most physicists think it to be. That is, an ether would be needed to provide a standard of handedness at each point of space, to tell Co^{60}, etc., which direction to decay into. But this is not

positively charged about nuclei What really exists here is a correlation between handedness and sign of charge. See SELECTION RULES (PHYSICS).

[C.J G]

Bibliography. T. D. Lee, Weak interactions and nonconservation of parity, *Science*, 127(3298): 569-573, 1958; T. D Lee and C. N. Yang, *Elementary Particles and Weak Interactions*, Brookhaven Natl. Lab BNL 443(T-91), 1957; P. Morrison, The overthrow of parity, *Sci. American*, 196(4): 45-53, 1957; C. N. Yang, Law of parity conservation and other symmetry laws, *Science*, 127(3298) 565-569, 1958

Parkinson's disease

A chronic, progressive disorder of the central nervous system, marked by slow movement and muscular rigidity, weakness, and tremor at rest. Parkinson's disease may result from certain forms of encephalitis, poisoning, head injury, syphilis, strokes, or commonly, arteriosclerosis. In its pure form it is the result of local impaired circulation that produces permanent damage to specific nerve cell groups in the brain. Such damage gives a typical clinical picture The patient is wide-eyed, staring, and has a masklike face. He walks with short shuffling steps, the body bent forward and the arms held stiffly at the sides Tremor is marked when a limb is at rest and may become so pronounced that involuntary "pill rolling" movements of the fingers and thumb appear. The tremor usually disappears upon movement, to be replaced by a cogwheel type of motion. Unless other brain areas have been affected by the precipitating agent, no mental changes are seen Excess salivation, fixation of the eyes, and speech impairment occur frequently Motor power is diminished but sensation is unimpaired. See CENTRAL NERVOUS SYSTEM, STRIATUM.

The disorder is slowly progressive and leads to increasing incapacity Treatment has been largely symptomatic and directed toward the relief of pain and cramps and the development of the best mental and occupational situation possible under the circumstances Since 1956, however, new developments in neurosurgery and chemotherapy have produced dramatic results in certain cases.

Relief of major symptoms has been obtained by producing lesions in certain brain centers, notably in the globus pallidus, one of the basal ganglia. The use of electrically induced lesions, as well as those resulting from ultrasonic vibrations and chemical injection, are being explored in selected cases with encouraging results [E.C.ST.]

Parrot

Any of many birds of the family Psittacidae, order Psittaciformes, a group which also includes the parakeets and macaws They are mostly tropical or subtropical, although one extinct species, the Carolina parakeet, *Conuropsis carolinensis*, formerly occurred in the southeastern and south central United States. The thick-billed parrot, *Rhynchopsitta pachyrhyncha*, of the pine forests of northern Mexico, is sometimes seen in Arizona.



The gray parrot, *Psittacus erithacus*, length to 12 in. (From E L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Many of the parrots are brightly colored birds, and they are caged as pets both because of their coloring and for the ability of some to talk, the best talkers are seldom brilliantly colored Parrots are expert climbers, having two toes forward and two back, but are clumsy walkers. Their food is seeds, nuts, and fruits. They usually nest in holes in trees See LOVEBIRD; PSITTACIFORMES. [J.D B]

Parsec

A unit of measure of astronomical distances One parsec is equivalent to 3.084×10^{13} kilometers, or 1.916×10^{13} miles. There are 3 26 light years in one parsec The parsec is defined as the distance at which the semimajor axis of Earth's orbit around the Sun (1 astronomical unit) subtends 1 second of arc Thus, because the angle is small

$$\frac{1 \text{ astronomical unit}}{1 \text{ parsec}} = 1 \text{ second} \approx \frac{1}{206{,}265}$$

A parsec is then 206,265 astronomical units; its accuracy depends on the precision with which the distance from Earth to Sun is measured At a distance of 1 parsec, the parallax is 1 second of arc See PARALLAX (ASTRONOMY) Stellar distances range from about 1 parsec for the nearest stars to several billion for the most distant known galaxy.

[J.L GR]

Intrinsic parity. The intrinsic parities of the particles composing a system must be multiplied by the orbital parity to yield the total parity. But the intrinsic parity of a *conserved* particle is irrelevant *and can be omitted. For if the particle is conserved* in a reaction, so is the contribution of its intrinsic parity to the total parity, so that its intrinsic parity is irrelevant to the balance of parity in the reaction. In fact, if a particle is conserved in all reactions, its intrinsic parity can never be determined. The photon is an unconserved particle; its intrinsic parity is odd. The parity of the π^0-meson (a pseudoscalar) is odd, so that to conserve parity it must be emitted by a nucleon into a P state. By charge independence, the charged π-meson must also be emitted in a P state; it is natural to call the parity of the charged π-meson odd also, which *amounts to defining the parity of the neutron and* proton to be the same. An electron by itself is conserved, but an electron plus a positron can annihilate. Thus the product of the parities of an electron and a positron must be well defined. According to the Dirac equation of relativistic quantum theory, the product of their parities is -1. The same result holds for any fermion particle-antiparticle pair Thus the parity of positronium is -1 times its orbital parity, that is, $-(-)^l$.

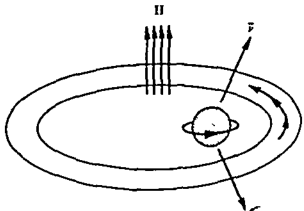
Spin and momentum correlations. The symmetry of the strong and electromagnetic interactions with respect to inversion implies statements about possible correlations of momenta and spins of the particles emitted as a result of such reactions. The principle is that the probability of a configuration of momenta and spins must be a scalar, in order that it not change under inversion of the coordinate system. Thus in a reaction yielding three particles with momenta $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, the angular distribution might be of the *form* $a + b\mathbf{p}_1 \cdot \mathbf{p}_2$ *but not* $a + b\mathbf{p}_1 \cdot \mathbf{p}_2 \times \mathbf{p}_3$, *for* under inversion the last term changes sign:

$$\mathbf{p}_1 \cdot \mathbf{p}_2 \times \mathbf{p}_3 \rightarrow (-\mathbf{p}_1) \cdot (-\mathbf{p}_2) \times (-\mathbf{p}_3) = -\mathbf{p}_1 \cdot \mathbf{p}_2 \times \mathbf{p}_3$$

This triple product is a pseudoscalar, and the description of the angular distribution would not be independent of the handedness of the coordinate system, because the coefficient b would appear to change sign. Orbital angular momentum, $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is a pseudovector, since under inversion $L \rightarrow +L$; the same must hold for spin angular momentum $\mathbf{S}$ Thus $\mathbf{S} \cdot \mathbf{p}$ is a pseudoscalar, and so such a term cannot occur in the angular distribution of a parity conserving process. This term, $\mathbf{S} \cdot \mathbf{p}$, in an angular distribution, would correlate a particle's spin with its momentum, that is, would imply a polarization in the momentum direction, or longitudinal polarization, which is accordingly absent in strong and electromagnetic reactions. Transverse polarizations, indicated *by terms such* as $\mathbf{S}_1 \cdot \mathbf{p}_1 \times \mathbf{p}_2$, are of course always possible.

Parity nonconservation. One of the selection rules which follows from parity conservation is the following The same spin zero boson cannot decay both into two π-mesons and three π-mesons, because these final states have opposite parities, ev and odd respectively. But the positive K-meson observed to do just this: it has both the $K_{\pi 2}$ a $K_{\pi 3}$ decay modes, and its spin is zero as deduc from the distribution of momenta in the $K_{\pi 3}$ mo Thus one concludes that parity is not conserved this decay. In 1956, T. D. Lee and C. N. Yang ma *the bold hypothesis that parity also is not co* served in β-decay. They reasoned that the mag tude of the β-decay coupling is about the same the coupling which leads to decay of the K-meso so these decay processes may be manifestations a single kind of coupling; and that there is a ve natural way to introduce parity nonconservati in β-decay, namely by assuming a restriction the possible states of the neutrino (two-compone theory). They pointed out that no β-decay expe ment had ever looked for the implied spin-mome tum correlations, which indicate parity nonconse vation; they urged that these correlations sought.

The first experiment to show parity nonconse vation in β-decay was done by Dr. T. Y. Wu collaboration with the physicists at the Nation Bureau of Standards. Here the spins of the β-acti nuclei cobalt-60 were polarized with a magnet field $\mathbf{H}$ at low temperature; the decay electro were observed to be emitted preferentially in d rections opposite to the direction of the Co^{60} spi (see illustration). Thus they found a $\mathbf{S}_{Co} \cdot \mathbf{p}_e$ co



Beta decay from polarized cobalt-60 nuclei. When the spin axes of the cobalt nuclei are not polarized, the preferential emission of the electrons and antineutrinos in the directions shown is not detectable.

relation; or in terms of macroscopic quantities, an $\mathbf{H} \cdot \mathbf{p}_e$ correlation. Later experiments have demonstrated the $\mathbf{S}_e \cdot \mathbf{p}_e$ correlation, that is, the longitudinal polarization of the electron. The magnitude of these correlations shows that the parity-nonconserving and parity-conserving parts of the β-interaction are of equal size, substantiating the two component neutrino theory.

It is now believed that parity conservation fails in all the weak decays, which includes all decays of the elementary particles (except *the* electromagnetic decays $\pi^0 \rightarrow 2\gamma$ and $\Sigma^0 \rightarrow \Lambda + \gamma$).

It was at first somewhat disconcerting to find parity not conserved, for that seemed to imply a handedness of space which would then not be the empty thing which (since the demise of the ether hypothesis) most physicists think it to be. That is, an ether would be needed to provide a standard of handedness at each point of space, to tell Co^{60}, etc., which direction to decay into. But this is not really the situation; the saving thing is that anti-Co^{60} decays in the opposite direction. Thus, after all, there is nothing intrinsically left handed about the world, just as there is nothing intrinsically positively charged about nuclei. What really exists here is a correlation between handedness and sign of charge. *See* SELECTION RULES (PHYSICS).

[C.J.G.]

Bibliography. T. D. Lee, Weak interactions and nonconservation of parity, *Science*, 127(3298): 569-573, 1958; T. D. Lee and C. N. Yang, *Elementary Particles and Weak Interactions*, Brookhaven Natl. Lab. BNL 443(T-91), 1957; P. Morrison, The overthrow of parity, *Sci. American*, 196(4): 45-53, 1957; C. N. Yang, Law of parity conservation and other symmetry laws, *Science*, 127(3298): 565-569, 1958.

Parkinson's disease

A chronic, progressive disorder of the central nervous system, marked by slow movement and muscular rigidity, weakness, and tremor at rest. Parkinson's disease may result from certain forms of encephalitis, poisoning, head injury, syphilis, strokes, or commonly, arteriosclerosis. In its pure form it is the result of local impaired circulation that produces permanent damage to specific nerve cell groups in the brain. Such damage gives a typical clinical picture. The patient is wide-eyed, staring, and has a masklike face. He walks with short shuffling steps, the body bent forward and the arms held stiffly at the sides. Tremor is marked when a limb is at rest and may become so pronounced that involuntary "*pill-rolling*" movements of the fingers and thumb appear. The tremor usually disappears upon movement, to be replaced by a cogwheel type of motion. Unless other brain areas have been affected by the precipitating agent, no mental changes are seen. Excess salivation, fixation of the eyes, and speech impairment occur frequently. Motor power is diminished but sensation is unimpaired. *See* CENTRAL NERVOUS SYSTEM; SYPHILIS.

The disorder is slowly progressive and leads to increasing incapacity. Treatment has been largely symptomatic and directed toward the relief of pain and cramps and the development of the best mental and occupational situation possible under the circumstances. Since 1956, however, new developments in neurosurgery and chemotherapy have produced dramatic results in certain cases.

Relief of major symptoms has been obtained by producing lesions in certain brain centers, notably in the globus pallidus, one of the basal ganglia. The use of electrically induced lesions, as well as those resulting from ultrasonic vibrations and chemical injection, are being explored in selected cases with encouraging results. [L.G.ST.]

Parrot

Any of many birds of the family Psittacidae, order Psittaciformes, a group which also includes the parakeets and macaws. They are mostly tropical or subtropical, although one extinct species, the Carolina parakeet, *Conuropsis carolinensis*, formerly occurred in the southeastern and south central United States. The thick-billed parrot, *Rhynchopsitta pachyrhyncha*, of the pine forests of northern Mexico, is sometimes seen in Arizona.



The gray parrot, *Psittacus erithacus*; length to 12 in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Many of the parrots are brightly colored birds, and they are caged as pets both because of their coloring and for the ability of some to talk; the best talkers are seldom brilliantly colored. Parrots are expert climbers, having two toes forward and two back, but are clumsy walkers. Their food is seeds, nuts, and fruits. They usually nest in holes in trees. *See* LOVEBIRD; PSITTACIFORMES. [J.D.B.]

Parsec

A unit of measure of astronomical distances. One parsec is equivalent to 3.084×10^{13} kilometers, or 1.916×10^{13} miles. There are 3.26 light years in one parsec. The parsec is defined as the distance at which the semimajor axis of Earth's orbit around the Sun (1 astronomical unit) subtends 1 second of arc. Thus, because the angle is small

$$\frac{1 \text{ astronomical unit}}{1 \text{ parsec}} = 1 \text{ second} = \frac{1}{206{,}265}$$

A parsec is then 206,265 astronomical units; its accuracy depends on the precision with which the distance from Earth to Sun is measured. At a distance of 1 parsec, the parallax is 1 second of arc. *See* PARALLAX (ASTRONOMY). Stellar distances range from about 1 parsec for the nearest stars to several billion for the most distant known galaxy.

[J.L.GR.]

Parsley

A biennial (*Petroselinum crispum*) of European origin belonging to the plant order Umbellales. Parsley is grown for its *foliage and is used to* garnish and flavor foods. It contains large quantities of vitamins A and C and has been grown for the past 2000 years or more. Two types, plain-leafed and curled, are grown for their foliage; Hamburg parsley (*P. crispum* var. *tuberosum*), also called turnip-rooted parsley, is grown for its edible parsniplike root. Propagation is by seed. Harvesting begins 70–80 days after planting for foliage varieties, and 90 days after planting for Hamburg parsley. Parsley is widely grown in the United States. *See* UMBELLALES; VEGETABLE GROWING.

[H.J.C.]

Parsnip

A hardy biennial (*Pastinaca sativa*) of Mediterranean origin belonging to the plant order Umbellales. The parsnip is grown for its thickened taproot and is used primarily as a cooked vegetable. Propagation is by seed; cultural practices are similar to those used for carrot, except that a longer growing season is required. Parsnip seed retains its viability only 1–2 years. Popular varieties are Hollow Crown and Model. Harvesting begins in late fall or early winter, usually 100–125 days after planting. Exposure of the mature roots to low temperatures, not necessarily below freezing, improves quality by favoring the conversion of starch to sugar. *See* CARROT, UMBELLALES; VEGETABLE GROWING.

[H.J.C.]

Parthenogenesis

A special type of sexual reproduction in which an egg develops without entrance of a sperm. It is common among rotifers, plant lice or aphids, thrips, in many ants, bees, wasps, and some crustaceans. Males are unknown in certain thrips and rotifers. Queen honey bees produce drones, or males, by parthenogenesis but also lay fertilized eggs that yield females, the workers, and queens. Aphids have successive generations of parthenogenetic females in spring and summer, then produce *both sexes by parthenogenesis*. These later mate, and the females lay fertilized eggs that hatch the following spring as females to resume parthenogenesis. *See* HOMOPTERA; HYMENOPTERA; ROTIFERA

[T.L.S.]

Partial differentiation

A mathematical operation performed on functions of more than one variable. In this article only two or three variables are considered; however, the principles apply to functions of n variables, for any positive integer $n > 1$. If $z = f(x,y)$, the partial derivative $\partial z/\partial x$ is defined as the derivative of $f(x,y)$ with respect to x, y *being regarded as fixed*. That is,

$$\frac{\partial z}{\partial x} = \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h}$$

Another notation for $\partial z/\partial x$ *is* $f_1(x,y)$. The other first partial derivative is $\partial z/\partial y$, also written $f_2(x,y)$. For values at particular points the notation is

$$\left(\frac{\partial z}{\partial x}\right)_{(a,b)} = f_1(a,b)$$

In the case of a function of three variables, $f(x,y,z)$, the expression is

$$\frac{\partial f}{\partial z} = f_3(x,y,z)$$

The second derivatives of $f(x,y)$ are

$$f_{11}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \qquad f_{12}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$
$$f_{21}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \qquad f_{22}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

It can happen that $f_{12}(x,y) \neq f_{21}(x,y)$, but this will not happen in common practice especially with elementary functions. If f_1, f_2, f_{12}, f_{21} are defined in a neighborhood of (a,b) and if f_{12}, f_{21} are continuous at (a,b), then $f_{12}(a,b) = f_{21}(a,b)$. There are more delicate theorems relating to this matter.

Differentials. The notions of a differential, and of the differentiability of a function, are fundamental in the theory of partial differentiation. For $f(x,y)$ to be differentiable is not the same as the requirement that $f_1(x,y)$ and $f_2(x,y)$ shall both exist; but *it is a more inclusive requirement. The geometric* meaning of f being differentiable at (a,b) is that the surface defined by $z = f(x,y)$ shall have a tangent plane not parallel to the z axis when $x = a$, $y = b$. *In analytic terms* the condition is that if

$$\epsilon = f(a+h,\ b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k$$

then

$$\lim_{(h,k)\to(0,0)} \frac{\epsilon}{|h|+|k|} = 0$$

When f is differentiable at (a,b) the expression $f_1(a,b)\,dx + f_2(a,b)\,dy$ is called the differential of f at (a,b) with independent increments dx and dy. It is a linear function of dx and dy, and among all linear functions $A\,dx + B\,dy$, it is the best approximation (in a definite sense of the word) to the expression

$$f(a+dx,\ b+dy) - f(a,b)$$

In the usual notation $z = f(x,y)$, one writes the differential, evaluated at (x,y), as

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

Here dx *and* dy *are independent variables and* dz is a dependent variable.

A sufficient condition that f be differentiable at (a,b) is that the partial derivatives f_1, f_2 be defined at all points near (a,b), and continuous at (a,b).

The chain rule. The prime importance of the differentiability concept lies in the fact that the differentiability property is needed in proving the chain rule for functions of several variables. This

rule asserts that a differentiable function of a differentiable function is differentiable, and the rule tells how to compute partial derivatives of the composite function. For example, if $x = f(s,t)$, $y = g(s,t)$, where f and g are differentiable, and if $z = F(x,y)$, where F is differentiable, then the composite function is $G(s,t) = F[f(s,t),g(s,t)]$, and its differential is

$$\frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy$$

where dx and dy, instead of being independent, are given by

$$dx = \frac{\partial f}{\partial s}ds + \frac{\partial f}{\partial t}dt \qquad dy = \frac{\partial g}{\partial s}ds + \frac{\partial g}{\partial t}dt$$

where ds and dt are independent. Then $z = G(s,t)$ is differentiable as a function of s and t, and

$$\frac{\partial G}{\partial s} = \frac{\partial F}{\partial x}\frac{\partial f}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial g}{\partial s} \qquad \frac{\partial G}{\partial t} = \frac{\partial F}{\partial x}\frac{\partial f}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial g}{\partial t}$$

These equations, expressing the formal part of the chain rule, are often written in the form

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

At one occurrence, the status of x and y is that of independent variables, as in $z = F(x,y)$, where they are called [illegible] of the second class.

The chain rule is valid for situations where there are any number of variables of the first class, and, quite independently, any number (the same or different) of the second class.

A typical use of the chain rule occurs when transformations are made on the variables in a problem. For example, one may switch from rectangular to polar coordinates. Then derivatives with respect to x, y, or both, must be converted into expressions involving derivatives with respect to r and θ. Transformations of variables are quite extensively used in studying partial differential equations.

Another interesting instance of the chain rule occurs in the so-called "particle-differentiation" in the flow of fluids. If $\rho = F(x,y,z,t)$ is the density at (x,y,z) in the fluid at time t, and if in a given motion one follows a certain selected particle, denoting the density of the fluid at this particle by $\rho = G(t)$, then

$$G'(t) = \frac{d\rho}{dt} = \frac{\partial \rho}{\partial x}\frac{dx}{dt} + \frac{\partial \rho}{\partial y}\frac{dy}{dt} + \frac{\partial \rho}{\partial z}\frac{dz}{dt} + \frac{\partial \rho}{\partial t}$$

Here ρ has two different meanings; on the left $\rho = G(t)$ and on the right $\rho = F(x,y,z,t)$. In dx/dt, dy/dt, dz/dt, the point (x,y,z) is the position of the particle being followed. Here the variables of the first class are x, y, z, t, and there is just one variable of the second class, namely t. The role of t also is different on the left and on the right in the equation.

Taylor developments. There is a Taylor's formula with remainder and a Taylor's series for functions of several variables. The easiest way to deal with these things is to think of them as being reduced back to the case of one variable by a device. If one wants to express $f(a + h, b + k)$ as a formula proceeding by terms of various degrees in h and k, consider

$$g(t) = f(a + th, b + tk)$$

develop $g(t)$ in powers of t and then set $t = 1$. The chain rule is needed to compute the derivatives of g. The general formula is

$$g^{(n)}(t) = \left[\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^n f(x,y)\right]_{x=a+th,\ y=b+tk}$$

Here a symbolic notation with rather evident meaning is used on the right.

Implicit functions. Suppose $F(x,y,z)$ is a function of three variables whose domain of definition is a certain collection of points (x,y,z) in space of three dimensions. As a general rule it will not be the case that the locus of points for which $F(x,y,z) = 0$ is the graph of an equation $z = f(x,y)$, where f is a single-valued function of two variables. But it may happen that, if (x_0,y_0,z_0) is a point of the locus $F(x,y,z) = 0$, there is a neighborhood of (x_0,y_0,z_0), consisting of all points inside a certain rectangular box centered at (x_0,y_0,z_0), such that the part of the locus $F(x,y,z) = 0$ inside this box is the graph of a function $z = f(x,y)$. There is a standard "implicit function theorem" which covers this situation. It states: Suppose F and its first partial derivatives F_1, F_2, F_3 are continuous throughout some specified neighborhood N of (x_0,y_0,z_0). Suppose also that $F(x_0,y_0,z_0) = 0$ and $F_3(x_0,y_0,z_0) \neq 0$. Then there exist certain positive constants a, b, c and a function f of x and y meeting all the following conditions. Let B denote the boxlike region composed of all (x,y,z) such that $|x-x_0| < a$, $|y-y_0| < b$, $|z-z_0| < c$, and let R denote the rectangle in the xy plane composed of all (x,y) such that $|x-x_0| < a$, $|y-y_0| < b$. The region B is contained in N, the function f is defined in R, and the graph of $z = f(x,y)$ is composed of precisely all the points in B at which $F(x,y,z) = 0$; f is continuous and has continuous first partial derivatives in R, given by

$$f_1(x,y) = -\frac{F_1(x,y,z)}{F_3(x,y,z)} \qquad f_2(x,y) = -\frac{F_2(x,y,z)}{F_3(x,y,z)}$$

where $z = f(x,y)$.

This theorem has two kinds of generalizations: one of the type in which F is a function of n variables and f is a function of n-1 variables, and the other of the type in which the locus $F(x,y,z) = 0$ is replaced by a locus defined by k equations in n variables ($n > k$), while the equation $z = f(x,y)$ is replaced by k equations involving k functions of n-k variables. Sample: $F(x,y,z,u,v) = 0$, $G(x,y,z,u,v) = 0$, $u = f(x,y,z)$, $v = g(x,y,z)$. Implicit function theorems of this second type are proved by mathe-

Parsley

A biennial (*Petroselinum crispum*) of European origin belonging to the plant order Umbellales. Parsley is grown for its foliage and is used to garnish and flavor foods. It contains large quantities of vitamins A and C and has been grown for the past 2000 years or more. Two types, plain-leafed and curled, are grown for their foliage; Hamburg parsley (*P. crispum* var. *tuberosum*), also called turnip-rooted parsley, is grown for its edible parsniplike root. Propagation is by seed. Harvesting begins 70–80 days after planting for foliage varieties, and 90 days after planting for Hamburg parsley. Parsley is widely grown in the United States. *See* UMBELLALES; VEGETABLE GROWING.

[H.J.C.]

Parsnip

A hardy biennial (*Pastinaca sativa*) of Mediterranean origin belonging to the plant order Umbellales. The parsnip is grown for its thickened taproot and is used primarily as a cooked vegetable. Propagation is by seed; cultural practices are similar to those used for carrot, except that a longer growing season is required. Parsnip seed retains its viability only 1–2 years. Popular varieties are Hollow Crown and Model. Harvesting begins in late fall or early winter, usually 100–125 days after planting. Exposure of the mature roots to low temperatures, not necessarily below freezing, improves quality by favoring the conversion of starch to sugar. *See* CARROT; UMBELLALES; VEGETABLE GROWING.

[H.J.C.]

Parthenogenesis

A special type of sexual reproduction in which an egg develops without entrance of a sperm. It is common among rotifers, plant lice or aphids, thrips, in many ants, bees, wasps, and some crustaceans. Males are unknown in certain thrips and rotifers. Queen honey bees produce drones, or males, by parthenogenesis but also lay fertilized eggs that yield females, the workers, and queens. Aphids have successive generations of parthenogenetic females in spring and summer, then produce both sexes by parthenogenesis. These later mate, and the females lay fertilized eggs that hatch the following spring as females to resume parthenogenesis. *See* HOMOPTERA; HYMENOPTERA; ROTIFERA

[T.I.S.]

Partial differentiation

A mathematical operation performed on functions of more than one variable. In this article only two or three variables are considered; however, the principles apply to functions of n variables, for any positive integer $n > 1$. If $z = f(x,y)$, the partial derivative $\partial z/\partial x$ is defined as the derivative of $f(x,y)$ with respect to x, y being regarded as fixed. That is,

$$\frac{\partial z}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x,y)}{h}$$

Another notation for $\partial z/\partial x$ is $f_1(x,y)$. The other first partial derivative is $\partial z/\partial y$, also written $f_2(x,y)$. For values at particular points the notation is

$$\left(\frac{\partial z}{\partial x}\right)_{(a,b)} = f_1(a,b)$$

In the case of a function of three variables, $f(x,y,z)$, the expression is

$$\frac{\partial f}{\partial z} = f_3(x,y,z)$$

The second derivatives of $f(x,y)$ are

$$f_{11}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \qquad f_{12}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$
$$f_{21}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \qquad f_{22}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

It can happen that $f_{12}(x,y) \neq f_{21}(x,y)$, but this will not happen in common practice especially with elementary functions. If f_1, f_2, f_{12}, f_{21} are defined in a neighborhood of (a,b) and if f_{12}, f_{21} are continuous at (a,b), then $f_{12}(a,b) = f_{21}(a,b)$. There are more delicate theorems relating to this matter.

Differentials. The notions of a differential, and of the differentiability of a function, are fundamental in the theory of partial differentiation. For $f(x,y)$ to be differentiable is not the same as the requirement that $f_1(x,y)$ and $f_2(x,y)$ shall both exist; but it is a more inclusive requirement. The geometric meaning of f being differentiable at (a,b) is that the surface defined by $z = f(x,y)$ shall have a tangent plane not parallel to the z axis when $x = a$, $y = b$. In analytic terms the condition is that if

$$\epsilon = f(a+h, b+k) - f(a,b) - f_1(a,b)h - f_2(a,b)k$$

then

$$\lim_{(h,k)\to(0,0)} \frac{\epsilon}{|h| + |k|} = 0$$

When f is differentiable at (a,b) the expression $f_1(a,b)\,dx + f_2(a,b)\,dy$ is called the differential of f at (a,b) with independent increments dx and dy. It is a linear function of dx and dy, and among all linear functions $A\,dx + B\,dy$, it is the best approximation (in a definite sense of the word) to the expression

$$f(a + dx,\ b + dy) - f(a,b)$$

In the usual notation $z = f(x,y)$, one writes the differential, evaluated at (x,y), as

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

Here dx and dy are independent variables and dz is a dependent variable.

A sufficient condition that f be differentiable at (a,b) is that the partial derivatives f_1, f_2 be defined at all points near (a,b), and continuous at (a,b).

The chain rule. The prime importance of the differentiability concept lies in the fact that the differentiability property is needed in proving the chain rule for functions of several variables. This

rule asserts that a differentiable function of a differentiable function is differentiable, and the rule tells how to compute partial derivatives of the composite function. For example, if $x = f(s,t)$, $y = g(s,t)$, where f and g are differentiable, and $z = F(x,y)$, where F is differentiable, then the composite function is $G(s,t) = F[f(s,t), g(s,t)]$, and its differential is

$$\frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy$$

where dx and dy, instead of being independent, are given by

$$dx = \frac{\partial f}{\partial s}ds + \frac{\partial f}{\partial t}dt \qquad dy = \frac{\partial g}{\partial s}ds + \frac{\partial g}{\partial t}dt$$

where ds and dt are independent. Then $z = G(s,t)$ is differentiable as a function of s and t, and

$$\frac{\partial G}{\partial s} = \frac{\partial F}{\partial x}\frac{\partial f}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial g}{\partial s} \qquad \frac{\partial G}{\partial t} = \frac{\partial F}{\partial x}\frac{\partial f}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial g}{\partial t}$$

These equations, expressing the formal part of the chain rule, are often written in the form

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

At one occurrence, the status of x and y is that of independent variables, as in $z = F(x,y)$, where they are called variables of the first class. But x, y also occur as dependent variables, depending on the independent variables s, t, which are called variables of the second class.

The chain rule is valid for situations where there are any number of variables of the first class and, quite independently, any number (the same or different) of the second class.

A typical use of the chain rule occurs when transformations are made on the variables in a problem. For example, one may switch from rectangular to polar coordinates. Then derivatives with respect to x, y, or both, must be converted into expressions involving derivatives with respect to r and θ. Transformations of variables are quite extensively used in studying partial differential equations.

Another interesting instance of the chain rule occurs in the so-called "material derivative" in the flow of fluids. If $p = F(x,y,z,t)$ is the density at (x,y,z) in the fluid at time t, and if in a given motion one follows a particular particle, noting the density of the fluid at the particle as $p = G(t)$, then

$$G'(t) = \frac{dp}{dt} = \frac{\partial p}{\partial x}\frac{dx}{dt} + \frac{\partial p}{\partial y}\frac{dy}{dt} + \frac{\partial p}{\partial z}\frac{dz}{dt} + \frac{\partial p}{\partial t}$$

Here p has two different meanings [illegible] $p = G(t)$ and [illegible] $p = F(x,y,z,t)$ [illegible] dy/dt, dz/dt [illegible] particle being followed [illegible] first class are [illegible] of the [illegible] in d[illegible] equations.

Taylor developments. There is a Taylor's formula with remainder and a Taylor's series for functions of several variables. The easiest way to deal with these things is to think of them as being reduced back to the case of one variable by a device. If one wants to express $f(a+h, b+k)$ as a formula proceeding by terms of various degrees in h and k, consider

$$g(t) = f(a+th, b+tk)$$

develop $g(t)$ in powers of t and then set $t = 1$. The chain rule is needed to compute the derivatives of g. The general formula is

$$g^{(n)}(t) = \left[\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^n f(x,y)\right]_{x=a+th,\, y=b+tk}$$

Here a symbolic notation with rather evident meaning is used on the right.

Implicit functions. Suppose $F(x,y,z)$ is a function of three variables whose domain of definition is a certain collection of points (x,y,z) in space of three dimensions. As a general rule it will not be the case that the locus of points for which $F(x,y,z) = 0$ is the graph of an equation $z = f(x,y)$, where f is a single-valued function of two variables. But it may happen that, if (x_0,y_0,z_0) is a point of the locus $F(x,y,z) = 0$, there is a neighborhood of (x_0,y_0,z_0), consisting of all points inside a certain rectangular box centered at (x_0,y_0,z_0), such that the part of the locus $F(x,y,z) = 0$ inside this box is the graph of a function $z = f(x,y)$. There is a standard "implicit function theorem" which covers this situation. It states: "Suppose F and its first partial derivatives F_1, F_2, F_3 are continuous throughout some specified neighborhood N of (x_0,y_0,z_0). Suppose also that $F(x_0,y_0,z_0) = 0$ and $F_3(x_0,y_0,z_0) \neq 0$. Then there exist certain positive constants a, b, c and a function f of x and y meeting all the following conditions: Let B denote the boxlike region composed of all (x,y,z) such that $|x - x_0| < a$, $|y - y_0| < b$, $|z - z_0| < c$, and let R denote the rectangle in the xy plane composed of all (x,y) such that $|x - x_0| < a$, $|y - y_0| < b$. The region B is contained in N; the function f is defined in R, and the graph of $z = f(x,y)$ is composed of precisely all the points in B at which $F(x,y,z) = 0$; f is continuous and has continuous first partial derivatives in R, given by

$$f_1(x,y) = -\frac{F_1(x,y,z)}{F_3(x,y,z)} \qquad f_2(x,y) = -\frac{F_2(x,y,z)}{F_3(x,y,z)}$$

where $z = f(x,y)$."

This theorem has far reaching kinds of generalizations; [illegible] F is a function of n variables and f is a function of $n - 1$ variables, and the [illegible] the locus $F(x,y,z) = 0$ [illegible] equations in [illegible] the equation $z = f(x,y)$ [illegible] functions of [illegible] $F(x,y,z,u,v) = 0$, $G(x,y,z,u,v) = 0$ [illegible] implicit function theorems [illegible] are proved [illegible] the conditions

in these theorems involve what are called *jacobian determinants*.

Jacobians. If $F_1, \ldots, F_k$ are k functions of $z_1, \ldots, z_k$, the determinant

$$J = \begin{vmatrix} \frac{\partial F_1}{\partial z_1} & \frac{\partial F_1}{\partial z_2} & \cdots & \frac{\partial F_1}{\partial z_k} \\ \frac{\partial F_2}{\partial z_1} & \cdots & & \frac{\partial F_2}{\partial z_k} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial F_k}{\partial z_1} & \cdots & & \frac{\partial F_k}{\partial z_k} \end{vmatrix}$$

is called the jacobian of $F_1, \ldots, F_k$ with respect to $z_1, \ldots, z_k$, and denoted by

$$J = \frac{\partial(F_1, \ldots, F_k)}{\partial(z_1, \ldots, z_k)}$$

Notice that the subscripts on the Fs are for distinguishing different functions, and do not indicate partial derivatives

The general implicit function theorem for a system of equations

$$F_1(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0, \ldots, \quad F_k(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0$$

guarantees a local solution of the form

$$z_1 = f_1(x_1, \ldots, x_r) \cdots z_k = f_k(x_1, \ldots, x_r)$$

near a set of values $x_i = a_i$, $z_j = b_j$ for which $F_1 = \ldots = F_k = 0$ and $J \neq 0$ This is on the assumption that the Fs have continuous first partial derivatives. The derivatives of the fs are given by the formulas

$$\frac{\partial f_i}{\partial x_p} = -\frac{J_{ip}}{J}$$

where J_{ip} is what J becomes when its ith column is replaced by $\partial F_1/\partial x_p, \ldots, \partial F_k/\partial x_p$.

If $u = f(x,y)$, $v = g(x,y)$ defines a one-to-one mapping of a region R_1 of the xy plane *onto* a region R_2 of the uv plane, and if it is known that f and g have continuous partial derivatives and the jacobian $J = [\partial(f,g)/\partial(x,y)]$ is never zero in R_1, then a simple closed curve C_1 in R_1 maps onto a simple closed curve C_2 in R_2 and, as a point P_1 goes counterclockwise around C_1, its image P_2 goes counterclockwise or clockwise around C_2 according as $J > 0$ or $J < 0$. Also, if A_1 and A_2 are the areas enclosed by C_1 and C_2 respectively, there is some point inside C_1 such that A_2/A_1 is the value of $|J|$ at that point. If a double integral with respect to u and v, over the region R_2, is converted into a double integral with respect to x and y over the region R_1, $du\,dv$ is replaced by $|J|dx\,dy$. These results generalize to the case of mappings in space of more than two dimensions. For example, in the passage from rectangular coordinates x, y, z to spherical polar coordinates r, θ, ϕ, by the equations $x = r \sin\phi \cos\theta$, $y = r \sin\phi \sin\theta$, $z = r \cos\phi$, the jacobian $\partial(x,y,z)/\partial(r,\theta,\phi)$ has the value $-r^2 \sin\phi$, and $dx\,dy\,dz$ is replaced by $r^2 \sin\phi\,dr\,d\theta\,d\phi$ in triple integrals

If the equations $u = f(x,y)$, $v = g(x,y)$ define a one-to-one mapping from the xy plane to the uv plane (in restricted regions), then

$$\frac{\partial(u,v)}{\partial(x,y)} = \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^{-1}$$

Functional dependence. If $f(x,y)$ and $g(x,y)$ are functionally dependent in a region R of the xy plane, then $[\partial(f,g)/\partial(x,y)] = 0$ in that region. An example of functional dependence would be this: $g(x,y) = [f(x,y)]^2 + \sin[f(x,y)]$. In general, f and g are called functionally dependent in R if there is some function F of u and v such that $F[f(x,y),g(x,y)] = 0$ at all points of R, and yet $F(u,v)$ is not zero throughout any two-dimensional portion of the uv plane. Conversely, if $[\partial(f,g)/\partial(x,y)] = 0$ at all points (x, y) in a neighborhood of (x_0,y_0), then usually f and g are functionally dependent in some (perhaps smaller) neighborhood of the point.

Homogeneous functions. One calls a function $F(x_1, \ldots, x_k)$ *positively homogeneous of degree n if* $F(tx_1, \ldots, tx_k) = t^n F(x_1, \ldots, x_k)$ *for all* $t > 0$ *and for all* $(x_1, \ldots, x_k)$ *in the domain* of definition of F. The index n need not be an integer. If F is differentiable and positively homogeneous of degree n, the Euler relation

$$x_1 \frac{\partial F}{\partial x_1} + \cdots + x_k \frac{\partial F}{\partial x_k} = nF(x_1, \ldots, x_k)$$

holds. Conversely, if F is differentiable in an open region which contains $(tx_1, \ldots, tx_k)$ for all $t > 0$, provided it contains $(x_1, \ldots, x_k)$, then the validity of Euler's relation in the region implies that F is positively homogeneous of degree n

Lagrange's method in extremal problems. If $F(x,y,z)$ is a differentiable function of three independent variables in an open region R of (x,y,z)-space, and if F reaches a relative maximum or minimum value at a point of R, then necessarily $(\partial F/\partial x) = (\partial F/\partial y) = (\partial F/\partial z) = 0$ there. Sufficient conditions, and tests for discrimination between maximum and minimum values, are sometimes stated in terms of second partial derivatives. Lagrange's method is concerned with the situation in which x, y, z are not independent, but are restricted by a side-condition $G(x,y,z) = 0$, where G is a specified function. Example: What is the maximum value of $x^2y^2z^2$ subject to the restriction $(x^2/25) + (y^2/16) + (z^2/9) - 1 = 0$? On the assumption that F and G have continuous first partial derivatives and that one never has $G = G_1 = G_2 = G_3 = 0$ at one point, the Lagrange procedure is to set $u = F + \lambda G$, where λ is a parameter. Then, among all the values of F attained for (x,y,z) such that $G(x,y,z) = 0$, if there is a maximum or minimum value, it will occur for an (x,y,z) point which satisfies the equations $F_i + \lambda G_i = 0$ $(i = 1, 2, 3)$ *and* $G = 0$, *for a certain value of* λ. These four equations can in theory be solved for x, y, z, λ, and the extreme value can be located. The method extends to other numbers of variables and to more than one side condition *See* CALCULUS, DIFFEREN-

TIAL AND INTEGRAL; DETERMINANT; DIFFERENTIATION, OPERATOR THEORY; PARAMETRIC EQUATION. [A.F.T.]

Bibliography: T. M. Apostol, *Mathematical Analysis*, 1957; R. Courant, *Differential and Integral Calculus*, 2 vols., 1936–1937; P. Franklin, *A Treatise on Advanced Calculus*, 1940; A. E. Taylor, *Advanced Calculus*, 1955.

Partial tone

A simple sinusoidal component of a complex tone. It may be part of a complex physical oscillation; alternatively, a partial tone (or partial) is a component of a sound sensation, distinguished as a simple tone that cannot be further analyzed by the ear and that contributes to the timbre of the complex sound. *See* TONE (MUSIC AND ACOUSTICS); *see also* HEARING.

The physical sound from a violin string, for example, is usually composed of a number of partials. Each of these partials of the air-borne sound results from vibration in a number of equal parts that takes place at the same time the string vibrates as a whole. Such characteristic vibrations by separate parts are also called partials; they are components of the complex motion of the string. Each such characteristic vibration pattern can also occur individually, however. Thus it seems preferable to call these characteristic motions modes of vibration rather [illegible]

[illegible]

If a string is bowed steadily, the frequencies of the partials of the resulting complex tone will be integral multiples of the lowest (fundamental) frequency, and the partials may properly be called harmonics. See HARMONIC (PERIODIC PHENOMENA). If, however, the same string is struck or plucked and then allowed to vibrate freely, the frequencies of the partials in the air borne sound and the frequencies of the corresponding modes of vibration are, in general, no longer exactly in the ratios of integers, and the partials and modes of vibration are inharmonic.

When a violinist produces a tone by touching a string lightly at its center while the string is bowed, both the sound in air and the vibration of the string are called harmonics; this can be misleading because it is now known that the frequency is not necessarily exactly twice that of the open string nor exactly equal to the frequency of the harmonic partial in the complex sound that results from steady bowing of the open string.

It is understandable on historical grounds how the terms partial, mode of vibration, harmonic, and overtone came to be used rather interchangeably. The purposes of musical science would be better served, however, by certain distinctions that could be attained by restricting the word harmonic to exact integral ratios of frequency, using mode of vibration as the name for the characteristic vibration, employing the word partial only to refer to a component of a complex sound, and avoiding entirely the use of the word overtone. *See* MUSICAL ACOUSTICS. [R.W.Y.]

Bibliography: H. L. F. Helmholtz, *On the Sensations of Tone*, A. J. Ellis, trans., 1875; R. W. Young, *Modes, nodes, and antinodes*, *Am. J. Phys.*, 20:177–183, 1952.

Particle

In classical mechanics, the term particle refers to a body having finite mass but negligible extension. A particle has inertia and possesses gravitational properties. Because of its negligible extension, forces acting on a particle cannot cause rotational acceleration; therefore, the motion of a particle is regarded as one of pure translation. An extended body is composed of particles. The translational motion of an extended body is equivalent to that of a single particle located at the center of mass of the body and having a mass equal to that of the entire body. *See* RIGID BODY; RIGID-BODY DYNAMICS.

The term particle is also used in physics as a synonym for "elementary particle." *See* ELEMENTARY PARTICLE. [D.W.I.]

Particle accelerator

A device which accelerates electrically charged atomic or subatomic particles to high energies. The particles can be electrons, protons, or ions such as deuterons, α-particles, and heavier ions.

Particle accelerators were conceived primarily as research tools in nuclear and particle physics. In this respect, they supplement by artificial radiations the natural sources available from radioactivity in the low energy region and cosmic rays at high energies. Cosmic rays are the only source of radiation in the extremely high-energy region; however, in the region where cosmic-ray and accelerator energies overlap, the advantages of intensity, collimation, and control make accelerators in general the preferred source of particles for research, except when problems of the origin and nature of cosmic radiation are an objective of the work. *See* COSMIC RAYS.

Accelerators capable of high intensity also have direct technical usefulness. Van de Graaff Generators, pulse transformer sets, cyclotrons, and electron linear accelerators are used to produce high volume rates of radiation for polymerization of plastics, *particularly in thin films, sterilization of* food, and radiation effect studies, such as radiation damage measurements of interest to reactor designers. The cyclotron is used extensively as a producer of neutron deficient isotopes that cannot be produced by reactor activation. Electrostatic generators, betatrons, and electron linear accelerators are useful as sources of x-rays of high penetrating power. X-ray and electron beams from Van de Graaff Generators and electron linear accelerators, and the external proton beams of synchrocyclotrons are used as radiation sources for localized cancer therapy. *See* IRRADIATION, ISOTOPIC, RADIOISOTOPE PRODUCTION.

Classification. The principles of operation of particle accelerators fall into two general classes,

in these theorems involve what are called jacobian determinants.

Jacobians. *If $F_1, \ldots, F_k$ are k functions of $z_1, \ldots, z_k$*, the determinant

$$J = \begin{vmatrix} \frac{\partial F_1}{\partial z_1} & \frac{\partial F_1}{\partial z_2} & \cdots & \frac{\partial F_1}{\partial z_k} \\ \frac{\partial F_2}{\partial z_1} & \cdots & & \frac{\partial F_2}{\partial z_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial F_k}{\partial z_1} & \cdots & & \frac{\partial F_k}{\partial z_k} \end{vmatrix}$$

is called the jacobian of $F_1, \ldots, F_k$ with respect to $z_1, \ldots, z_k$, and denoted by

$$J = \frac{\partial(F_1, \ldots, F_k)}{\partial(z_1, \ldots, z_k)}$$

Notice that the subscripts on the Fs are for distinguishing different functions, and do not indicate partial derivatives.

The general implicit function theorem for a system of equations

$$F_1(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0, \ldots, F_k(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0$$

guarantees a local solution of the form

$$z_1 = f_1(x_1, \ldots, x_r) \cdots z_k = f_k(x_1, \ldots, x_r)$$

near a set of values $x_i = a_i$, $z_j = b_j$ for which $F_1 = \ldots = F_k = 0$ and $J \neq 0$. This is on the assumption that the Fs have continuous first partial derivatives. The derivatives of the fs are given by the formulas

$$\frac{\partial f_i}{\partial x_p} = -\frac{J_{ip}}{J}$$

where J_{ip} is what J becomes when its ith column is replaced by $\partial F_1/\partial x_p, \ldots, \partial F_k/\partial x_p$.

If $u = f(x,y)$, $v = g(x,y)$ defines a one-to-one mapping of a region R_1 of the xy plane onto a region R_2 of the uv plane, and if it is known that f and g have continuous partial derivatives and the jacobian $J = [\partial(f,g)/\partial(x,y)]$ is never zero in R_1, then a simple closed curve C_1 in R_1 maps onto a simple closed curve C_2 in R_2 and, as a point P_1 goes counterclockwise around C_1, its image P_2 goes counterclockwise or clockwise around C_2 according as $J > 0$ *or* $J < 0$. *Also, if A_1 and A_2 are the* areas enclosed by C_1 and C_2 respectively, there is some point inside C_1 such that A_2/A_1 is the value of $|J|$ at that point. If a double integral with respect to u and v, over the region R_2, is converted into a double integral *with respect to x and y* over the region R_1, $du\,dv$ is replaced by $|J|dx\,dy$. These results generalize to the case of mappings in space of more than two dimensions. For example, in the passage from rectangular coordinates x, y, z to spherical polar coordinates r, θ, ϕ, by the equations $x = r \sin\phi \cos\theta$, $y = r \sin\phi \sin\theta$, $z = r \cos\phi$, the jacobian $\partial(x,y,z)/\partial(r,\theta,\phi)$ has the value $-r^2 \sin\phi$, and $dx\,dy\,dz$ is replaced by $r^2 \sin\phi\,dr\,d\theta\,d\phi$ in triple integrals.

If the equations $u = f(x,y)$, $v = g(x,y)$ (a one-to-one mapping from the xy plane to t *plane (in restricted regions), then*

$$\frac{\partial(u,v)}{\partial(x,y)} = \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^{-1}$$

Functional dependence. *If $f(x,y)$ and $g(x,y$* functionally dependent in a region R of th plane, then $[\partial(f,g)/\partial(x,y)] = 0$ in that regio example of functional dependence would be $g(x,y) = [f(x,y)]^2 + \sin[f(x,y)]$. In gener and g are called functionally dependent in there is some function F of u and v such $F[f(x,y),g(x,y)] = 0$ at all points of R, an $F(u,v)$ is not zero throughout any two-dimens portion of the uv plane. Conversely, if $[\partial(f,$ $\partial(x,y)] = 0$ at all points (x, y) in a neighbor of (x_0,y_0), then usually f and g are functio dependent in some (perhaps smaller) neig hood of the point.

Homogeneous functions. *One calls a fun* $F(x_1, \ldots, x_k)$ positively homogeneous of gree n if $F(tx_1, \ldots, tx_k) = t^nF(x_1, \ldots, x_k)$ all $t > 0$ and for all $(x_1, \ldots, x_k)$ in the do of definition of F. The index n need not be a teger. *If F is differentiable and positively hom neous of degree n*, the Euler relation

$$x_1\frac{\partial F}{\partial x_1} + \cdots + x_k\frac{\partial F}{\partial x_k} = nF(x_1, \ldots, x_k)$$

holds. Conversely, if F is differentiable in an o region which contains $(tx_1, \ldots, tx_k)$ *for all* 0, *provided it contains* $(x_1, \ldots, x_k)$, then the lidity of Euler's relation in the region implies t *F is positively homogeneous of degree n.*

Lagrange's method in extremal problems. $F(x,y,z)$ is a differentiable function of three in pendent variables in an open region R of $(x,y,$ space, and if F reaches a relative maximum minimum value at a point of R, then necessa $(\partial F/\partial x) = (\partial F/\partial y) = (\partial F/\partial z) = 0$ there. Su cient conditions, and tests for discrimination tween maximum and minimum values, are so times stated in terms of second partial derivati Lagrange's method is concerned with the situat in which x, y, z are not independent, but are stricted by a side-condition $G(x,y,z) = 0$, where is a specified function Example· What is the ma *mum value of* $x^2y^2z^2$ *subject to the restricti* $(x^2/25) + (y^2/16) + (z^2/9) - 1 = 0$? On the a sumption that F and G have continuous first part derivatives and that one never has $G = G_1 = G_2$ $G_3 = 0$ at one point, the Lagrange procedure is *set $u = F + \lambda G$, where λ is a parameter* The among all the values of F attained for (x,y,z) su that $G(x,y,z) = 0$, if there is a maximum or mi mum value, it will occur for an (x,y,z) point whi satisfies the equations $F_i + \lambda G_i = 0$ $(i = 1, 2, 3$ and $G = 0$, *for a certain value of λ. These fo* equations can in theory be solved for x, y, z, λ and the extreme value can be located The meth extends to other numbers of variables and to mo than one side condition. *See* CALCULUS, DIFFERE

tial and integral; Determinant; Differentiation; Operator theory; Parametric equation. [A.E.T.]

Bibliography: T. M. Apostol, *Mathematical Analysis*, 1957; R. Courant, *Differential and Integral Calculus*, 2 vols., 1936–1937; P. Franklin, *A Treatise on Advanced Calculus*, 1940; A. E. Taylor, *Advanced Calculus*, 1955.

Partial tone

A simple sinusoidal component of a complex tone. It can be part of a complex physical oscillation; alternatively, a partial tone (or partial) is a component of a sound sensation, distinguished as a simple tone that cannot be further analyzed by the ear and that contributes to the timbre of the complex sound. *See* Tone (music and acoustics); *see also* Timbre.

The physical sound from a violin string, for example, is usually composed of a number of partials. Each of these partials of the air borne sound results from vibration in a number of equal parts that takes *place* at the same time the string vibrates as a whole. Such characteristic vibrations by separate parts are also called partials; they are components of the complex motion of the string. Each such characteristic vibration pattern can also occur individually, *however*. Thus it seems preferable to call these characteristic motions modes of vibration rather than partials, to make it clear that they can exist independently without being parts of a complex motion. *See* Mode of vibration; Vibration.

If a string is bowed *steadily*, the frequencies of the partials of the resulting complex tone will be integral multiples of *the lowest* (fundamental) frequency, and the partials may properly be called harmonics. *See* Harmonic (periodic phenomena). If, however, the same *string is* struck or plucked and then allowed to vibrate freely, the frequencies of the partials in the air borne sound and the frequencies of the corresponding modes of vibration are, in general, no longer exactly in the ratios of integers, and the partials and modes of vibration are inharmonic.

When a violinist produces a tone by touching a string lightly at its center while the string is bowed, both the sound *in air* and the vibration of the string are called harmonics; this can be misleading because it is now *known that the* frequency is not necessarily exactly twice that of the open string nor exactly equal to *the* frequency of the harmonic partial in the complex sound that results from steady bowing of the open *string*.

It is understandable on historical grounds how the terms partial, mode of vibration, harmonic, and overtone came to be used rather *interchangeably*. The purposes of musical science would be better served, however, by certain distinc*tions* that could be attained by restricting the word harmonic to exact integral ratios of frequency, using mode of vibration as the name for the charac*teristic* vibration, employing the word partial only to refer to a component of a complex sound, and avoiding entirely the use of the word overtone. *See* Musical acoustics. [R.W.Y.]

Bibliography: H. L. F. Helmholtz, *On the Sensations of Tone*, A. J. Ellis, trans., 1875; R. W. Young, Modes, nodes, and antinodes, *Am. J. Phys.*, 20:177–183, 1952.

Particle

In classical mechanics, the term particle refers to a body having finite mass but negligible extension. A particle has inertia and possesses gravitational properties. Because of its negligible extension, forces acting on a particle cannot cause rotational acceleration; therefore, the motion of a particle is regarded as one of pure translation. An extended body is composed of particles. The translational motion of an extended body is equivalent to that of a single particle located at the center of mass of the body and having a mass equal to that of the entire body. *See* Rigid body; Rigid-body dynamics.

The term particle is also used in physics as a synonym for "elementary particle." *See* Elementary particle. [D.W.]

Particle accelerator

A device which accelerates electrically charged atomic or subatomic particles to high energies. The particles can be electrons, protons, or ions such as deuterons, α-particles, and heavier ions.

Particle accelerators were conceived primarily as research tools in nuclear and particle physics. In this respect, they supplement by artificial radiations the natural sources available from radioactivity in the low-energy region and cosmic rays at high energies. Cosmic rays are the only source of radiation in the extremely high-energy region. However, in the region where cosmic-ray and accelerator energies overlap, the advantages of intensity, collimation, and control make accelerators in general the preferred source of particles for research, except when problems of the origin and nature of cosmic radiation are an objective of the work. *See* Cosmic rays.

Accelerators capable of high intensity also have direct technical usefulness. Van de Graaff Generators, pulse transformer sets, cyclotrons, and electron linear accelerators are used to produce high volume rates of radiation for polymerization of plastics, particularly in thin films, sterilization of food, and radiation-effect studies, such as radiation damage measurements of interest to reactor designers. The cyclotron is used extensively as a producer of neutron-deficient isotopes that cannot be produced by reactor activation. Electrostatic generators, betatrons, and electron linear accelerators are useful as sources of x-rays of high penetrating power. X ray and electron beams from Van de Graaff Generators and electron linear accelerators, and the external proton beams of synchrocyclotrons are used as radiation sources for localized cancer therapy. *See* Irradiation, isotopic; Radioisotope production.

Classification. The principles of operation of particle accelerators fall into two general classes,

in these theorems involve what are called jacobian determinants.

Jacobians. If $F_1, \ldots, F_k$ are k functions of $z_1, \ldots, z_k$, the determinant

$$J = \begin{vmatrix} \frac{\partial F_1}{\partial z_1} & \frac{\partial F_1}{\partial z_2} & \cdots & \frac{\partial F_1}{\partial z_k} \\ \frac{\partial F_2}{\partial z_1} & \cdot & \cdot & \frac{\partial F_2}{\partial z_k} \\ \cdots & \cdots & \cdots & \\ \frac{\partial F_k}{\partial z_1} & \cdots & & \frac{\partial F_k}{\partial z_k} \end{vmatrix}$$

is called the jacobian of $F_1, \ldots, F_k$ with respect to $z_1, \ldots, z_k$, and denoted by

$$J = \frac{\partial(F_1, \ldots, F_k)}{\partial(z_1, \ldots, z_k)}$$

Notice that the subscripts on the Fs are for distinguishing different functions, and do not indicate partial derivatives.

The general implicit function theorem for a system of equations

$$F_1(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0, \ldots, F_k(x_1, \ldots, x_r, z_1, \ldots, z_k) = 0$$

guarantees a local solution of the form

$$z_1 = f_1(x_1, \ldots, x_r) \cdots z_k = f_k(x_1, \ldots, x_r)$$

near a set of values $x_i = a_i$, $z_j = b_j$ for which $F_1 = \ldots = F_k = 0$ and $J \neq 0$ This is on the assumption that the Fs have continuous first partial derivatives. The derivatives of the fs are given by the formulas

$$\frac{\partial f_i}{\partial x_p} = -\frac{J_{ip}}{J}$$

where J_{ip} is what J becomes when its ith column is replaced by $\partial F_1/\partial x_p, \ldots, \partial F_k/\partial x_p$.

If $u = f(x,y)$, $v = g(x,y)$ defines a one-to-one mapping of a region R_1 of the xy plane onto a region R_2 of the uv plane, and if it is known that f and g have continuous partial derivatives and the jacobian $J = [\partial(f,g)/\partial(x,y)]$ is never zero in R_1, then a simple closed curve C_1 in R_1 maps onto a simple closed curve C_2 in R_2 and, as a point P_1 goes counterclockwise around C_1, its image P_2 goes counterclockwise or clockwise around C_2 according as $J > 0$ or $J < 0$ Also, if A_1 and A_2 are the areas enclosed by C_1 and C_2 respectively, there is some point inside C_1 such that A_2/A_1 is the value of $|J|$ at that point. If a double integral with respect to u and v, over the region R_2, is converted into a double integral with respect to x and y over the region R_1, $du\,dv$ is replaced by $|J|dx\,dy$. These results generalize to the case of mappings in space of more than two dimensions For example, in the passage from rectangular coordinates x, y, z to spherical polar coordinates r, θ, ϕ, by the equations $x = r \sin\phi \cos\theta$, $y = r \sin\phi \sin\theta$, $z = r \cos\phi$, the jacobian $\partial(x,y,z)/\partial(r,\theta,\phi)$ has the value $-r^2 \sin\phi$, and $dx\,dy\,dz$ is replaced by $r^2 \sin\phi\, dr\, d\theta\, d\phi$ in triple integrals.

If the equations $u = f(x,y)$, $v = g(x,y)$ [...] a one-to-one mapping from the xy plane to t[...] plane (in restricted regions), then

$$\frac{\partial(u,v)}{\partial(x,y)} = \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^{-1}$$

Functional dependence. If $f(x,y)$ and $g(x,y$[...] functionally dependent in a region R of t[...] plane, then $[\partial(f,g)/\partial(x,y)] = 0$ in that regio[...] example of functional dependence would be $g(x,y) = [f(x,y)]^2 + \sin[f(x,y)]$. In gene[...] and g are called functionally dependent in [...] there is some function F of u and v such [...] $F[f(x,y),g(x,y)] = 0$ at all points of R, an[...] $F(u,v)$ is not zero throughout any two-dimens[...] portion of the uv plane. Conversely, if $[\partial(f$[...] $\partial(x,y)] = 0$ at all points (x, y) in a neighbor[...] of (x_0,y_0), then usually f and g are functio[...] dependent in some (perhaps smaller) neig[...] hood of the point.

Homogeneous functions. One calls a fun[...] $F(x_1, \ldots, x_k)$ positively homogeneous of [...] gree n if $F(tx_1, \ldots, tx_k) = t^n F(x_1, \ldots, x_k)$ [...] all $t > 0$ and for all $(x_1, \ldots, x_k)$ in the do[...] of definition of F. The index n need not be a[...] teger. If F is differentiable and positively hom[...] neous of degree n, the Euler relation

$$x_1 \frac{\partial F}{\partial x_1} + \quad + x_k \frac{\partial F}{\partial x_k} = nF(x_1, \ldots, x_k)$$

holds. Conversely, if F is differentiable in an o[...] region which contains $(tx_1, \ldots, tx_k)$ for all [...] 0, provided it contains $(x_1, \ldots, x_k)$, then the [...] lidity of Euler's relation in the region implies [...] F is positively homogeneous of degree n.

Lagrange's method in extremal problems. [...] $F(x,y,z)$ is a differentiable function of three i[...] pendent variables in an open region R of $(x,y$[...] space, and if F reaches a relative maximum [...] minimum value at a point of R, then necessa[...] $(\partial F/\partial x) = (\partial F/\partial y) = (\partial F/\partial z) = 0$ there. S[...] cient conditions, and tests for discrimination [...] tween maximum and minimum values, are so[...] times stated in terms of second partial derivativ[...] Lagrange's method is concerned with the situati[...] in which x, y, z are not independent, but are [...] stricted by a side-condition $G(x,y,z) = 0$, where [...] is a specified function. Example: What is the ma[...] mum value of $x^2y^2z^2$ subject to the restricti[...] $(x^2/25) + (y^2/16) + (z^2/9) - 1 = 0$? On the [...] sumption that F and G have continuous first part[...] derivatives and that one never has $G = G_1 = G_2$ [...] $G_3 = 0$ at one point, the Lagrange procedure is [...] set $u = F + \lambda G$, where λ is a parameter. The [...] among all the values of F attained for (x,y,z) su[...] that $G(x,y,z) = 0$, if there is a maximum or mi[...] mum value, it will occur for an (x,y,z) point whi[...] satisfies the equations $F_i + \lambda G_i = 0$ ($i = 1, 2, 3$[...] and $G = 0$, for a certain value of λ. These fo[...] equations can in theory be solved for x, y, z, [...] and the extreme value can be located The metho[...] extends to other numbers of variables and to mo[...] than one side condition. *See* CALCULUS, DIFFERE[...]

Table 1. Operating characteristics of particle accelerators

Accelerator type	Particle accelerated	Energy range	Beam current (average peak)	Duty cycle	Energy spectrum*	Beam geometry	Development status (1960)
Electrostatic accelerators							
Cockcroft-Walton generator	p, d, α, e	4 Mev	1 ma, 1 ma	Continuous	~0.01%	Small focal spot	4 Mev operating
Impulse generator	p, d, α, e	2 Mev	Small 10 ma	Small	~1%	Fairly large focal spot	2 Mev operating
Van de Graaff Generator	p, d, α, e	10 Mev	1 ma, 10 ma	Continuous	~0.01%	Small focal spot	8.5 Mev operating, 10-Mev design (tandem)
Resonant and pulse-transformer sets	p, e	2 Mev	100 ma, 1 amp	<½	Broad	Fairly large focal spot	2 Mev operating
Time-varying field accelerators							
Circular magnetic types (radio-frequency resonance accelerators)							
Cyclotron	p, d, α, heavy ion	25 Mev (d)	250 ma, 250 ma	Continuous	~1%	Internal target or external beam of fair collimation at lower intensity external neutrons external meson beams	22 Mev (d) operating
Synchrocyclotron	p, d, α	900 Mev	1 μa, 100 μa	$<10^{-3}$	0.1%	Internal target or external beam of fair collimation at lower intensity external neutrons external meson beams	720-Mev operating
Electron synchrotron	e	2 Bev	Very small to 5 μa, depends on duty cycle	$<10^{-3}$	0.1%	External x-ray beam	1.1-Bev operating
Proton synchrotron	p	1-12 Bev	0.05 μa depends on duty cycle	$<10^{-3}$	0.1%	Internal targets external beam of fair collimation at lower intensity external neutrons external secondary particle beams	10-Bev operating
Alternating-gradient synchrotron	p, e	10–70 Bev (p), 0.5–7 Bev (e)	0.05 μa (p), 1 μa (e) depends on duty cycle	$<10^{-3}$	0.1%	(p) internal targets external beam of fair colli mation at lower intensity external neutrons, external second ary particle beams (e) exter nal γ ray beam external electron beam proposed	29-Bev operating (p) 70 Bev design (p) 6-Bev construction (e)
Fixed-field, alternating-gradient synchrotron	p, e	To 25 Bev	0.1 ma, depends on duty cycle	$<10^{-3}$	0.1%	Internal targets and external secondary-particle beams possibility of colliding beams	Model and design studies
Circular magnetic type (induction accelerator)							
Betatron	e	10–500 Mev	0.1 μa, 10 μa	10^{-3}	0.05%	External x-ray beam external electron beam of excellent collimation of lower intensity	340-Mev operating
Linear accelerators							
Heavy-particle linear accelerator	p, d, α, heavy ions	1 Bev	10 μa, 1 ma	10^{-3}	0.5%	Well-collimated and well-focused external beam	70-Mev (p) operating high-energy design (p) 10 Mev/nucleon (heavy-ion)
Electron linear accelerator	e	6 Mev to 45 Bev	60 μa 200 ma	10^{-3}	2%	Well-collimated and well focused external beam	700-Mev operating 45 Bev design

* Spread in energy of beam expressed as a percentage of total energy of beam that is, 1% for the cyclotron means for a 1-Mev beam a spread in energy of 0.01 Mev

electrostatic accelerators and accelerators employing time-varying electric or magnetic fields. Electrostatic accelerators give the particle its energy by letting the particle travel through an evacuated tube containing the ion or electron source on one end and the target at the other; the source and target are maintained at a difference of electric potential by various means. The maximum energy which can be attained is limited by the techniques available in producing and insulating an electrostatic voltage. This limitation does not apply to accelerators employing fields which vary with time. The action of time-varying fields falls into two fundamental classes: (1) devices such as the cyclotron, which utilize means of having the particle traverse a time-varying potential difference many times; and (2) devices using nonstatic electric fields for acceleration, that is, electric fields associated with changing magnetic fields. Devices of the first kind are based on the principle that, in contrast to motion in a constant field, the energy gain of a charged particle having described a closed orbit in an electrostatic (that is, derivable from a scalar potential) but time-varying field is not zero. Acceleration of this type is also called transit-time acceleration since the existence of an energy gain in a static field is made possible by the time lags due to finite particle speeds. Devices of the second type make use of Faraday's law, the relation which gives the total energy gain of a charge describing a closed orbit in terms of the time rate of change of the magnetic flux through the orbit. See FARADAY'S LAW OF INDUCTION.

Accelerators can also be classified as circular or linear. Circular machines employ an electromagnet which produces a magnetic guide-field to bend the particle orbits into circles or circular arcs joined by straight sections. The guide-field is in general separate from the fields involved in the acceleration. Linear accelerators accelerate particles in nearly straight-line orbits; formally this term includes the electrostatic accelerators, although it is rarely so used.

It is customary to measure the energy of the particles accelerated in units of electron volts (ev); millions of electron volts (10^6 ev = 1 Mev), or billions (10^9 ev = 1 Bev, or 1 GeV in Europe). A particle is defined to have a kinetic energy of T electron volts if it has the same energy as a particle carrying one electronic charge which has been accelerated in an electrostatic accelerator operating at a voltage $V = T$. See ELECTRON VOLT.

All practical accelerators designed through 1959 have employed the basic principles outlined: acceleration of individual charged particles in externally produced large-scale electromagnetic fields. Suggestions have frequently been made to design an accelerator based on cooperative action of many atomic particles upon one (for example, inducing many particles in a discharge to communicate their energy to one member), but these ideas have not yet led to practical designs. They are attractive in principle since they would free accelerator design from limitations imposed by the properties of materials, such as the saturation of iron in the case of magnetic fields and the vacuum breakdown limit in the case of electric fields.

Characteristics. In addition to the basic distinctions of operating principles, particle accelerators differ in terms of many parameters affecting their usefulness for various purposes. The important accelerator characteristics are: kind of particle accelerated, particle energy, beam intensity (number of particles accelerated per second), duty cycle (fraction of the time the beam is available in case the machine is pulsed) and pulse length and repetition rate, beam geometry (configuration of the primary beam and of charged and neutral secondary beams as to accessibility, angular divergence, and beam diameter), beam energy spectrum and beam purity, and ease and range of control of energy and intensity. All the accelerators described in this article differ in these operating characteristics. It is for this reason that particle accelerators have been constructed in such a wide variety of designs; the various applications demand different characteristics.

Table 1 lists the more common types of accelerators and the characteristics important to their use. The energy and intensity ranges and other parameters given are approximate ranges of values attained with existing machines or design parameters of planned devices. They are not intended to represent fundamental limits of performance of each machine, although, as will be evident from the more detailed discussion of each machine, some of the attained values are close to the practical maxima.

ELECTROSTATIC ACCELERATORS

All electrostatic accelerators employ an evacuated discharge tube which can be a continuous insulator, but more commonly is a series of metallic electrodes spaced by insulators. Such a discharge tube, and in particular the gaps between the electrodes, will have a focusing action known as electrostatic focusing resulting from an electrostatic lens, or more precisely, second-order electrostatic focusing. The action can be illustrated by considering a gap between cylindrical electrodes (Fig. 1). Consider a charged particle traveling from a to b, let the direction of the fields be such that the particle is accelerated. At A, the particle will receive an impulse toward the axis, while when crossing a line of force at B it will receive an impulse away from the axis. If the particle velocity were constant, the total impulse received would be exactly zero. However, because of the acceleration, the particle spends less time in the diverging field region than in the converging field region, and hence, a net focusing impulse results. Note that this impulse depends quadratically on the field (hence the name, second-order focusing): the field produces the radial forces and also the velocity change which gives the net action. The reader can go through the argument for a decelerating lens and find that the focusing action is still converging. It is fairly easy to write a differential equation for the orbits or use graphical [illegible]

Table 1. Operating characteristics of particle accelerators

Accelerator type	Particle accelerated	Energy range	Beam current (average, peak)	Duty cycle	Energy spectrum*	Beam geometry	Development status (1960)
Electrostatic accelerators							
Cockcroft-Walton generator	p, d, α, e	4 Mev	1 ma, 1 ma	Continuous	~0.01%	Small focal spot	4 Mev operating
Impulse generator	p, d, α, e	2 Mev	Small, 10 ma	Small	~1%	Fairly large focal spot	2 Mev operating
Van de Graaff Generator	p, d, α, e	10 Mev	1 ma, 10 ma	Continuous	~0.01%	Small focal spot	8.5-Mev operating, 10 Mev design (tandem)
Resonant and pulse-transformer sets	p, e	2 Mev	100 ma, 3 amp	<½	Broad	Fairly large focal spot	2 Mev operating
Time-varying field accelerators							
Circular magnetic types (radio-frequency resonance accelerators)							
Cyclotron	p, d, α, heavy ion	25 Mev (d)	250 μa, 250 μa	Continuous	~1%	Internal target or external beam of fair collimation at lower intensity, external neutrons, external meson beams	22 Mev (d) operating
Synchrocyclotron	p, d, α	900 Mev	1 μa, 100 μa	<10[illegible]	0.1%	Internal target or external beam of fair collimation at lower intensity, external neutrons, external meson beams	720-Mev operating
Electron synchrotron	e	2 Bev	Very small to 5 μa, depends on duty cycle	<10[illegible]	0.1%	External x-ray beam	1.1-Bev operating
Proton synchrotron	p	1-12 Bev	0.05 μa, depends on duty cycle	<10[illegible]	0.1%	Internal targets, external beam of fair collimation at lower intensity, external neutrons, external secondary-particle beams	10-Bev operating
Alternating-gradient synchrotron	p, e	10-70 Bev (p), 0.5-7 Bev (e)	0.05 μa (p), 1 μa (e), depends on duty cycle	<10[illegible]	0.1%	(p) internal targets, external beam of fair collimation at lower intensity, external neutrons, external secondary-particle beams; (e) external γ-ray beam, external electron beam proposed	29-Bev operating (p), 70-Bev design (p), 6-Bev construction (e)
Fixed-field, alternating-gradient synchrotron	p, e	To 25 Bev	0.1 ma, depends on duty cycle	<10[illegible]	0.1%	Internal targets and external secondary-particle beams, possibility of colliding beams	Model and design studies
Circular magnetic type (induction accelerator)							
Betatron	e	10-500 Mev	0.1 μa, 10 μa	10[illegible]	0.05%	External x-ray beam, external electron beam of excellent collimation of lower intensity	340-Mev operating
Linear accelerators							
Heavy particle linear accelerator	p, d, α, heavy ions	1 Bev	10 μa, 1 ma	10[illegible]	0.5%	Well-collimated and well focused external beam	70-Mev (p) operating, high-energy design (p), 10 Mev/nucleon (heavy-ion)
Electron linear accelerator	e	6 Mev to 45 Bev	60 μa, 200 ma	10[illegible]	2%	Well-collimated and well-focused external beam	700-Mev operating, 45-Bev design

*Spread in energy of beam expressed as a percentage of total energy of beam, that is, 1% for the cyclotron means for a 1-Mev beam a spread in energy of 0.01 Mev.

electrostatic accelerators and accelerators employing time-varying electric or magnetic fields. Electrostatic accelerators give the particle its energy by letting the particle travel through an evacuated tube containing the ion or electron source on one end and the target at the other; the source and target are maintained at a difference of electric potential by various means. The maximum energy which can be attained is limited by the techniques available in producing and insulating an electrostatic voltage. This limitation does not apply to accelerators employing fields which vary with time. The action of time-varying fields falls into two fundamental classes: (1) devices such as the cyclotron, which utilize means of having the particle traverse a time-varying potential difference many times; and (2) devices using nonstatic electric fields for acceleration, that is, electric fields associated with changing magnetic fields Devices of the first kind are based on the principle that, in contrast to motion in a constant field, the energy *gain of a charged particle having described* a closed orbit in an electrostatic (that is, derivable from a scalar potential) but time-varying field is not zero. Acceleration of this type is also called transit-time acceleration since the existence of an energy gain in a static field is made possible by the time lags due to finite particle speeds. Devices of the second type make use of Faraday's law, the relation which gives the total energy gain of a charge describing a closed orbit in terms of the time rate of change of the magnetic flux through the orbit. *See* FARADAY'S LAW OF INDUCTION.

Accelerators can also be classified as circular or linear. Circular machines employ an electromagnet which produces a magnetic guide-field to bend the particle orbits into circles or circular arcs joined by straight sections. The guide-field is in general separate from the fields involved in the acceleration. Linear accelerators accelerate particles in nearly straight-line orbits; formally this term includes the electrostatic accelerators, although it is rarely so used.

It is customary to measure the energy of the particles accelerated in units of electron volts (ev); millions of electron volts (10^6 ev = 1 Mev), or billions (10^9 ev = 1 Bev, or 1 GeV in Europe). A particle is defined to have a kinetic energy of T electron volts if it has the same energy as a particle carrying one electronic charge which has been accelerated in an electrostatic accelerator operating at a voltage $V = T$. *See* ELECTRON VOLT

All practical accelerators designed through 1959 have employed the basic principles outlined: acceleration of individual charged particles in externally produced large-scale electromagnetic fields. Suggestions have frequently been made to design an accelerator based on cooperative action of many atomic particles upon one (for example, inducing many particles in a discharge to communicate their energy to one member), but these ideas have not yet led to practical designs. They are attractive in principle since they would free accelerator design from limitations imposed by the properties of materials, such as the saturation of iron in the case of magnetic fields and the vacuum breakdown limit in the case of electric fields.

Characteristics. In addition to the basic distinctions of operating principles, particle accelerators differ in terms of many parameters affecting their usefulness for various purposes. The important accelerator characteristics are: kind of particle accelerated, particle energy, beam intensity (number of particles accelerated per second), duty cycle (fraction of the time the beam is available in case the machine is pulsed) and pulse length and repetition rate, beam geometry (configuration of the primary beam and of charged and neutral secondary beams as to accessibility, angular divergence, and beam diameter), beam energy spectrum and beam purity, and ease and range of control of energy and intensity. All the accelerators described in this article differ in these operating characteristics. It is for this reason that particle accelerators have been constructed in such a wide variety of designs; the various applications demand different characteristics.

Table 1 lists the more common types of accelerators and the characteristics important to their use The energy and intensity ranges and other parameters given are approximate ranges of values attained with existing machines or design parameters of planned devices. They are not intended to represent fundamental limits of performance of each machine, although, as will be evident from the more detailed discussion of each machine, some of the attained values are close to the practical maxima.

ELECTROSTATIC ACCELERATORS

All electrostatic accelerators employ an evacuated discharge tube which can be a continuous insulator, but more commonly is a series of metallic electrodes spaced by insulators. Such a discharge tube, and in particular the gaps between the electrodes, will have a focusing action known as electrostatic focusing resulting from an electrostatic lens, or more precisely, second-order electrostatic focusing. The action can be illustrated by considering a gap between cylindrical electrodes (Fig. 1) Consider a charged particle traveling from a to b; let the direction of the fields be such that the particle is accelerated At A, the particle will receive an impulse toward the axis, while when crossing a line of force at B it will receive an impulse away from the axis If the particle velocity were constant, the total impulse received would be exactly zero However, because of the acceleration, the particle spends less time in the diverging field region than in the converging field region, and hence, a net focusing impulse results. Note that this impulse depends quadratically on the *field* (hence the name, second-order focusing) · the field produces the radial forces and also the velocity change which gives the net action. The reader can go through the argument for a decelerating lens and find that the focusing action is still converging. It is fairly easy to write a differential equation for the orbits or use graphical methods; an approxi-

Table 1 Operating characteristics of particle accelerators

Accelerator type	Particle accelerated	Energy range	Beam current (average, peak)	Duty cycle	Energy spectrum*	Beam geometry	Development status (1960)
Electrostatic accelerators							
Cockcroft-Walton generator	p, d, α, e	4 Mev	1 ma, 1 ma	Continuous	~0 01%	Small focal spot	4 Mev operating
Impulse generator	p, d, α, e	2 Mev	Small, 10 ma	Small	~1%	Fairly large focal spot	2 Mev operating
Van de Graaff Generator	p, d, α, e	10 Mev	1 ma, 10 ma	Continuous	~0 01%	Small focal spot	7 5-Mev operating, 10-Mev design (tandem)
Resonant and pulse-transformer sets	p, e	2 Mev	100 ma, 1 amp	<½	Broad	Fairly large focal spot	2 Mev operating
Time-varying field accelerators							
Circular magnetic types (radio-frequency resonance accelerators)							
Cyclotron	p, d, α heavy ion	25 Mev (d)	250 ma, 250 ma	Continuous	~1%	Internal target or external beam of fair collimation at lower intensity, external neutrons, external meson beams	22 Mev (d) operating
Synchrocyclotron	p, d, α	900 Mev	1 μa, 100 μa	<10[illegible]	0 1%	Internal target or external beam of fair collimation at lower intensity, external neutrons external meson beams	720-Mev operating
Electron synchrotron	e	2 Bev	Very small to 5μa depends on duty cycle	<10[illegible]	0 1%	External x-ray beam	1 1-Bev operating
Proton synchrotron	p	1-12 Bev	0 05 μa, depends on duty cycle	<10[illegible]	0 1%	Internal targets external beam of fair collimation at lower intensity, external neutrons external secondary particle beams	10-Bev operating
Alternating-gradient synchrotron	p, e	10-70 Bev (p) 0 5-7 Bev (e)	0 05 μa (p) 1 μa (e) depends on duty cycle	<10[illegible]	0 1%	(p) internal targets external beam of fair collimation at lower intensity external neutrons, external secondary-particle beams, (e) external γ-ray beam external electron beam proposed	29-Bev operating (p), 70-Bev design (p) 6-Bev construction (e)
Fixed field, alternating-gradient synchrotron	p, e	To 25 Bev	0 1 ma depends on duty cycle	<10[illegible]	0 1%	Internal targets and external secondary-particle beams possibility of colliding beams	Model and design studies
Circular magnetic type (induction accelerator)							
Betatron	e	10-500 Mev	0 1 μa, 10 μa	10[illegible]	0 05%	External x-ray beam external electron beam of excellent collimation of lower intensity	340-Mev operating
Linear accelerators							
Heavy-particle linear accelerator	p, d, α, heavy ions	1 Bev	10 μa, 1 ma	10[illegible]	0 5%	Well-collimated and well-focused external beam	70-Mev (p) operating, high-energy design (p), 10 Mev/nucleon (heavy-ion)
Electron linear accelerator	e	6 Mev to 45 Bev	60 μa 200 ma	10[illegible]	2%	Well-collimated and well focused external beam	700-Mev operating 45-Bev design

* Spread in energy of beam expressed as a percentage of total energy of beam, that is 1% for the cyclotron means for a 1-Mev beam a spread in energy of 0 01 Mev

electrostatic accelerators and accelerators employing time-varying electric or magnetic fields. Electrostatic accelerators give the particle its energy by letting the particle travel through an evacuated tube containing the ion or electron source on one end and the target at the other; the source and target are maintained at a difference of electric potential by various means. The maximum energy which can be attained is limited by the techniques available in producing and insulating an electrostatic voltage. This limitation does not apply to accelerators employing fields which vary with time. The action of time-varying fields falls into two fundamental classes: (1) devices such as the cyclotron, which utilize means of having the particle traverse a time-varying potential difference many times; and (2) devices using nonstatic electric fields for acceleration, that is, electric fields associated with changing magnetic fields. Devices of the first kind are based on the principle that, in contrast to motion in a constant field, the energy gain of a charged particle having described a closed orbit in an electrostatic (that is, derivable from a scalar potential) but time-varying field is not zero. Acceleration of this type is also called transit-time acceleration since the existence of an energy gain in a static field is made possible by the time lags due to finite particle speeds. Devices of the second type make use of Faraday's law, the relation which gives the total energy gain of a charge describing a closed orbit in terms of the time rate of change of the magnetic flux through the orbit. *See* FARADAY'S LAW OF INDUCTION.

Accelerators can also be classified as circular or linear. Circular machines employ an electromagnet which produces a magnetic guide-field to bend the particle orbits into circles or circular arcs joined by straight sections. The guide-field is in general separate from the fields involved in the acceleration. Linear accelerators accelerate particles in nearly straight-line orbits; formally this term includes the electrostatic accelerators, although it is rarely so used

It is customary to measure the energy of the particles accelerated in units of electron volts (ev); millions of electron volts (10^6 ev = 1 Mev), or billions (10^9 ev = 1 Bev, or 1 GeV in Europe). A particle is defined to have a kinetic energy of T electron volts if it has the same energy as a particle carrying one electronic charge which has been accelerated in an electrostatic accelerator operating at a voltage $V = T$. *See* ELECTRON VOLT.

All practical accelerators designed through 1959 have employed the basic principles outlined: acceleration of individual charged particles in externally produced large-scale electromagnetic fields. Suggestions have frequently been made to design an accelerator based on cooperative action of many atomic particles upon one (for example, inducing many particles in a discharge to communicate their energy to one member), but these ideas have not yet led to practical designs. They are attractive in principle since they would free accelerator design from limitations imposed by the properties of materials, such as the saturation of iron in the case of magnetic fields and the vacuum breakdown limit in the case of electric fields.

Characteristics. In addition to the basic distinctions of operating principles, particle accelerators differ in terms of many parameters affecting their usefulness for various purposes. The important accelerator characteristics are: kind of particle accelerated, particle energy, beam intensity (number of particles accelerated per second), duty cycle (fraction of the time the beam is available in case the machine is pulsed) and pulse length and repetition rate, beam geometry (configuration of the primary beam and of charged and neutral secondary beams as to accessibility, angular divergence, and beam diameter), beam energy spectrum and beam purity, and ease and range of control of energy and intensity. All the accelerators described in this article differ in these operating characteristics. It is for this reason that particle accelerators have been constructed in such a wide variety of designs; the various applications demand different characteristics.

Table 1 lists the more common types of accelerators and the characteristics important to their use. The energy and intensity ranges and other parameters given are approximate ranges of values attained with existing machines or design parameters of planned devices. They are not intended to represent fundamental limits of performance of each machine, although, as will be evident from the more detailed discussion of each machine, some of the attained values are close to the practical maxima.

ELECTROSTATIC ACCELERATORS

All electrostatic accelerators employ an evacuated discharge tube which can be a continuous insulator, but more commonly is a series of metallic electrodes spaced by insulators. Such a discharge tube, and in particular the gaps between the electrodes, will have a focusing action known as electrostatic focusing resulting from an electrostatic lens, or more precisely, second-order electrostatic focusing. The action can be illustrated by considering a gap between cylindrical electrodes (Fig. 1). Consider a charged particle traveling from a to b, let the direction of the fields be such that the particle is accelerated. At A, the particle will receive an impulse toward the axis, while when crossing a line of force at B it will receive an impulse away from the axis. If the particle velocity were constant, the total impulse received would be exactly zero However, because of the acceleration, the particle spends less time in the diverging field region than in the converging field region, and hence, a net focusing impulse results. Note that this impulse depends quadratically on the field (hence the name, second-order focusing): the field produces the radial forces and also the velocity change which gives the net action. The reader can go through the argument for a decelerating lens and find that the focusing action is still converging It is fairly easy to write a differential equation for the orbits or use graphical methods; an approxi-

and

$$\omega_C = \frac{BZe}{\gamma m_0 c} = \frac{1}{\gamma}\omega_0 \quad (10)$$

The quantity ω_C is known as the cyclotron frequency and is characterized by the fact that for particle velocities $v << c$ (that is, $\gamma \approx 1$),

$$\omega_C \rightarrow \omega_0 = BZe/m_0 c \quad (11)$$

which is a quantity independent of the particle energy at a given value of the field.

It is customary to represent the radial fall-off of the field by an index n defined by

$$n = -\frac{r_0}{B_0}\frac{\partial B}{\partial r} \quad (12)$$

where r_0 and B_0 are the values of the orbit radius r and the induction B about which the particles oscillate.

It can be shown that, as the result of the z axis restoring action (as shown in Fig. 3) and the geometry of the orbits in their plane, the particles will oscillate about the equilibrium orbit, where the frequencies for small oscillations are given by

$$\omega_z = \omega_C\sqrt{n} \quad (13)$$

along the field lines, and by

$$\omega_r = (1-n)^{1/2}\omega_C \quad (14)$$

perpendicular to the field lines, respectively. These frequencies are known as the betatron frequencies, and the oscillations as betatron oscillations. As the result of the radial oscillation, the center of the [orbi]t will *precess* about the center of symmetry [wit]h a frequency

$$\omega_p = 1-(1-n)^{1/2} \quad (15)$$

Strong focusing. When n becomes larger than [unit]y in a magnetic field of circular symmetry, the [radi]al motion becomes unstable, that is, the par[tic]les *spiral out.* Hence, in the weak-focusing sys[te]m, n is restricted to the interval $0 < n < 1$. Gen[era]lly, other considerations narrow the permissible [ra]nge of n further. If $n >> 1$, the particles are [ve]rtically very stable and horizontally unstable; if [n] $<< -1$, they are vertically unstable and horizontally very stable. It is known that lenses of [e]qual converging and diverging strength, but separated by a small distance, form a converging com[bi]nation regardless of which lens comes first (Fig. 4). Hence, if n alternates between large positive and negative values, that is, if the radial field gradient is reversed as a function of angle around the orbit, a net focusing action can result both radially and along the field lines. This is known as alternating gradient (AG) focusing. There are additional restrictions on the frequency of alternation of gradient and on n which must be met if stability is to result; these restrictions are too complex to discuss here. Focusing obtained by alternating focusing and defocusing (of which AG focusing is an example) is often called strong focusing; it is characterized by the fact that the frequencies ω_z and ω_r can become much larger than the cyclotron frequency ω_C, in contrast to the weak-focusing case expressed by Eqs. (13) and (14). This causes certain complications: if ω_z or ω_r coincides with harmonics of ω_C, energy of the main orbital motion can be fed into the radial and axial oscillations leading to instability; hence, such integral relations must be avoided or be passed very rapidly in the case of time-varying fields.

Cyclotron resonance, phase stability. Equation (11) shows that in a constant magnetic field and for nonrelativistic motion ($v << c$), the time to complete a circular orbit is independent of particle energy. Hence, if an alternating voltage of angular frequency $\omega_C \approx \omega_0$ is applied across a radial gap located in the magnetic field, the particles will cross this gap always at the same phase of the alternating voltage (Fig. 5); if a particle is started at a phase suitable for acceleration, it will spiral

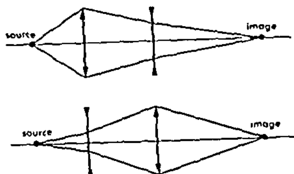


Fig 4 Model of strong focusing, a combination of a converging and a diverging lens yields a net converging system; $n >> 1$ is converging for axial motion and diverging for radial oscillations, $n << -1$ is diverging for axial motion and converging for radial motion

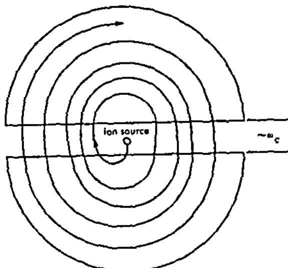


Fig 5. Principle of the cyclotron. The ions are formed in the ion source and are drawn out by one of the D's. The particles are bent into circular orbits, the period of the orbit is equal to the period ($2\pi/\omega_C$) of the applied D voltage. The particle will thus cross the gap between the D's at constant phase

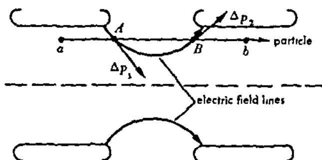


Fig. 1. Focusing action of an electrostatic lens Δp_1 is the inward momentum acquired in the first half of the lens and Δp_2 is the outward momentum in the second half (assuming the field to be accelerating). Focusing action results since $\Delta p_1 > \Delta p_2$ due to the shorter time spent by the particle in the second half of the lens

mate formula for the focal length f of a lens, such as the one shown, for a particle of kinetic energy T and charge e in an electric field $E(z)$ is

$$\frac{1}{f} = \frac{3}{16}\int \frac{(eE)^2}{T^2}\,dz \qquad (1)$$

For additional information, *see* ELECTROSTATIC LENS; *see also* ELECTRON MOTION IN VACUUM.

In practice, only the first few gaps of an accelerator column have a major lens effect; their adjustment is critical, since it is desirable that the beam be neither underfocused (radially expanding) nor overfocused (crossing over in the column, leading to excessive final angular divergence). The important types of electrostatic accelerators are discussed in detail in separate articles, *see* COCKCROFT-WALTON ACCELERATOR; RESONANCE TRANSFORMER; VAN DE GRAAFF GENERATOR

CIRCULAR ACCELERATORS

Particle orbits. Circular accelerators utilize a magnetic field of induction B to bend charged-particle orbits and confine the extent of particle motion (*see* INDUCTION, MAGNETIC; MAGNETIC FIELD) It is known from electromagnetic theory that a magnetic field exerts a force on a moving charge perpendicular to the direction of the field and to the particle velocity; as the result of this force, the orbit of a particle carrying Z electronic charges e has a radius of curvature r given by

$$1/r = BZe/cp_{\perp} \qquad (2a)$$

(Gaussian, cgs units; omit the factor c if mks units are used), or in units frequently used in particle physics,

$$1/r \cong 300BZ/cp_{\perp} \qquad (2b)$$

where B is in gauss, cp in electron volts, and r in centimeters. Here $p_{\perp}$ is the component of the particle momentum p perpendicular to the magnetic field lines. Equations (2a) and (2b) are correct for particles moving at relativistic velocities, provided the relativistic equations

$$cp = \beta\gamma m_0c^2 \qquad (3)$$

$$E = \gamma m_0c^2$$
$$\gamma = (1-\beta^2)^{-1/2}$$
$$\beta = v/c$$
$$E^2 = (cp)^2 + (m_0c^2)^2$$

are used to relate the mechanical quantities m rest mass, c = velocity of light, E = total ene (including rest energy), v = particle velocity. I uniform magnetic field, a particle will thus scribe a helix of radius given by Eqs. (2a) (2b) and pitch given by the initial conditi (Fig. 2). *See* RELATIVISTIC ELECTRODYNAMI RELATIVISTIC MECHANICS; RELATIVITY.

Weak focusing. A uniform magnetic field w in general, not confine the particles after t execute a few turns. A guide-field which falls radially in magnitude (Fig. 3) will produce restoring force which causes the orbits to oscill about the midplane of the magnetic field Fr Eqs. (2a) and (3), it is found that the time T the corresponding angular frequency

$$\omega_C = 2\pi/T$$

to describe a complete circular orbit are given by

$$T = \frac{2\pi r}{v} = \frac{2\pi\gamma m_0 c}{BZe}$$

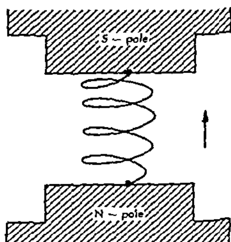


Fig 2. Motion in a uniform magnetic field of indu tion B The orbit is a helix of parameters as given Eqs. (2a) and (2b).

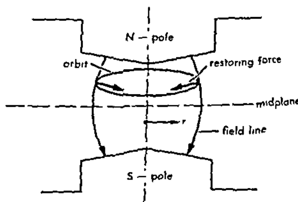


Fig. 3. Motion in a magnetic field of circular symme try with radial fall-off.

acceleration in relatively few turns; this requires very large rf power sources (100 kw or more for the larger cyclotrons) and also makes the support of the D-structure a difficult problem, since insulators become hot through rf losses. In most cyclotrons, the D's are supported on "D-stems" which are electrically a quarter wavelength long when considered as transmission lines with capacitative loading to represent the D's (Fig. 7). This permits the D-stems to be grounded on the end away from the D. The high powered oscillator couples power at the region of high rf magnetic field near the shorted end. The oscillator can be independently tuned, or (more commonly) the resonance frequency of the D-structure itself can be the frequency-defining element. The ion source consists generally of an arc discharge along the axis of the cyclotron from pole to pole; gas is fed into the arc to supply the ions; the vacuum tank is pumped at high speed to maintain an adequate vacuum (see Fig. 8).

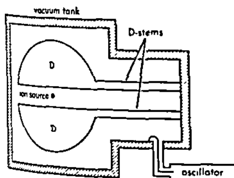


Fig. 7 Schematic representation of cyclotron D-structure. The entire structure constitutes a network resonant at the frequency ω_0.

Synchrocyclotron. As is seen from Table 2, the synchrocyclotron employs the phase-stability principle, and uses a programmed frequency of the rf system to guide the particles to higher energy and larger radii. Since the motion is stable, it is permissible to have the particles execute very many turns, and thus, relatively low D voltages are required; in fact, the D voltage level is determined by the need to draw the ions out of the source and the desired time program of the beam. The oscillator is frequency modulated either by a rotating condenser or by a large electrically driven tuning fork.

The synchrocyclotron resembles the ordinary cyclotron in appearance, but because of the lower D-voltage requirement, the magnet gap is generally smaller in proportion. Usually, only one D is used, the opposite electrode is a grounded dummy D.

The beam of the synchrocyclotron is pulsed in accordance with the time cycle of the frequency modulation, hence, its average beam current is much lower than that of the ordinary cyclotron. The beam can strike an internal target placed at a suitable radius, or the beam can be extracted.

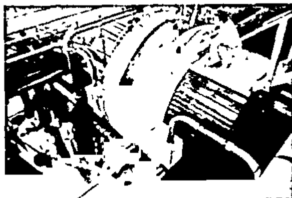


Fig. 8 The 90-in. cyclotron at the University of California Radiation Laboratory, Livermore. This is an unusual design employing a C-shaped magnet and permitting a variable energy. (*University of California*)

The most successful extraction method consists of exciting large-amplitude radial oscillations in the beam by a number of magnetic irregularities placed at suitable angular positions in the magnet; these oscillations increase the turn spacing to such an extent that a large fraction of the beam can be caught in a magnetic shield which "pipes" the particles out of the magnet (see Fig. 9).

Electron synchrotron. As is shown in Table 2, the electron synchrotron makes use of the phase-stability principle; the particle energy follows an increasing magnetic field at a constant accelerating frequency.

Electrons attain a velocity close to that of light at very moderate energies, the radius of the orbits will change only by about 2% as the particle is accelerated from 2 Mev to very high energies. Hence, contrary to the design for cyclotron magnets, the magnet can be in the form of a ring to produce a magnetic field only in an annular volume. The cost and weight of a magnet are related to the magnetic



Fig. 9 View of the vacuum tank and upper magnet coil of the 184-in. synchrocyclotron of the University of California Radiation Laboratory, Berkeley. Pumps are in the foreground, target-handling equipment is at the lower left and the rf oscillator housing is at the right-hand edge. Note the magnet return yoke and the shielding at the top of the picture. (*University of California*)

out to the field boundaries as its energy increases. This is the principle of the cyclotron: the oscillator driving the electrodes (D's) must be adjusted to *the frequency ω_C, given by Eq. (11); ω_C is defined* only by the kind of ion to be accelerated and the magnetic field.

The cyclotron-resonance concept is only approximate for two reasons: *(a)* the relativistic factor γ in Eq. (3) representing the mass increase has been ignored, and *(b)* to obtain axial focusing, a small n-value is required so that the field B in Eqs. (2*a*) and (2*b*) is not constant, but falls off with r. Both of these effects throw the particles out of step with the accelerating voltage in the same direction. Thus, the ordinary cyclotron is limited in the number of turns during which acceleration can occur, and hence the required D voltage becomes very high at large energies

What happens to a particle not exactly in step with the frequency of the accelerating system in the general case when these approximate assumptions are not made? Consider a given radius at which the magnetic field is given and where the D's are excited at a given frequency. Consider a particle which crosses the gap in Fig. 5 at such a phase that it is accelerated, but let the field be decreasing during the time of passage. Assume that this particle has a momentum less than that required by Eq (11) for the conditions in question. This means that it will move to a smaller radius, and hence (for $n > 0$) to a stronger field This effect, combined with the relativistic factor γ in Eq. (11), will make ω_C too large; hence, the next time around, the particle will arrive earlier and will thus gain more energy when passing the gap. Thus, the particle whose energy was too low for the conditions of D frequency and magnetic field at a given radius will gain extra energy, and vice versa. The particle energy will, in this way, oscillate about the correct energy. This condition is known as phase stability, and in the case of circular accelerators (but not for linear accelerators), occurs for those particles which cross the accelerating gap at a time when the accelerating field is decreasing (Fig. 6). The phase stability action will lock the particles into orbits such that their energy $\gamma m_0 c^2$ and radius r are defined in terms of the applied frequency ω_C of the accelerating system and of the magnetic field B (or, $\omega_0 = BZe/m_0c$). Specifically, ω_C and ω_0 define r and γ by the relations, derived from Eqs. (2*a*), (2*b*), and (11),

$$\frac{r}{c} = \left(\frac{1}{\omega_C^2} - \frac{1}{\omega_0^2}\right)^{1/2} \qquad (16)$$

and

$$\gamma = \left[1 + \left(\frac{\omega_0 r}{c}\right)^2\right]^{1/2} = \frac{\omega_0}{\omega_C} \qquad (17)$$

It follows that it is possible to accelerate the particle by changing either the frequency ω_C or the magnetic induction B slowly (adiabatically), provided the radio-frequency (rf) power is sufficient to provide the increasing energy per turn. This is the action of the various types of "synchro-" machines, as shown in Table 2 If one of the parameters is changed at a certain rate, then the particle whose energy exactly keeps step with the change will ride at the synchronous-phase angle φ_S shown in Fig 6; the other particles will execute phase oscillations about φ_S. In general, the amplitudes of the phase and betatron oscillations will decrease with increasing particle energy.

Table 2. Programming of parameters of various types of circular accelerators

Accelerator	Frequency of accelerating system, ω_C	Magnetic guide-field, ω_0	Radius, r
Synchrocyclotron*	Decreasing	Constant	Increasin
Electron synchrotron, $v \approx c$	Constant	Increasing	Constant
Proton synchrotron (including AGS)	Increasing	Increasing	Constant
Ordinary cyclotron	Constant	Constant	Increasin

* Sometimes referred to as synchrophasotron in the Russian literature.

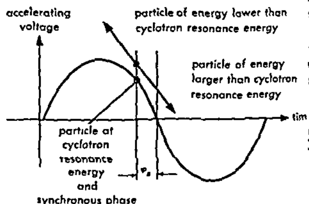


Fig. 6. Principle of phase stability in circular accelerators. The diagram shows the accelerating voltage as a function of time A particle of correct energy and radius in the magnetic field to be in cyclotron resonance will cross the accelerating gap at a phase angle φ_S (synchronous phase) as shown; this phase is defined such that $\cos \varphi_S$ is the ratio of energy gain per turn required by the accelerating program divided by the maximum energy gain possible. Particles of energy below resonance energy will tend to move toward phases of higher energy gain and vice versa; this action results in phase stability.

Fixed-field cyclotron. Most fixed-field cyclotrons accelerate deuterons (D^+), α-particles (He^{++}), and molecular hydrogen ions (H_2^+), since these have nearly the same charge-to-mass ratio, and thus, *nearly the same cyclotron frequency* Fixed-field cyclotrons have been built with magnet pole diameters up to 90 in., giving energies of 30 Mev to deuterons, 60 Mev to α-particles, and 15 Mev to protons (half the energy of the H_2^+ ion). The magnet is usually a structure containing the two circular poles, two coils close to the poles, and two magnetic return circuits. The accelerating structure (D's) is contained in a vacuum chamber inside the poles. As mentioned before, it is necessary to operate the D's at very high voltages to achieve the

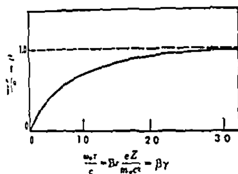


Fig. 11. Plot of particle velocity β (in units of the velocity of light c) against particle momentum $\beta\gamma$ (in units of rest energy divided by c). The particle velocity is a measure of the required frequency ω_c of the accelerating system, and the momentum is a measure of the magnetic field. The curve thus represents the required program of magnetic field and frequency.

There are, however, many additional practical differences. Proton synchrotrons usually operate at a lower rate of rise of magnetic field, and hence, the magnet lamination can be thicker without excessive eddy-current effects. Also, the vacuum chamber can be fabricated of thin stainless steel or other low-conductivity metal. Injection is accomplished by a separate electrostatic accelerator or proton linear accelerator and a pulsed electric inflecting system. Proton synchrotrons have been constructed in the form of four quadrants separated by straight sections. The inflection electrodes, the rf accelerating cavities, and other apparatus are located in the field free straight sections. Figure 12 shows a typical arrangement.

The rf system and its control present special problems since the accelerating cavity (or electrode) must be modulated over a wide frequency range in accordance with the program shown in Fig. 11. To give the accelerating structure enough bandwidth to handle the frequency range would require excessive power. The best solution appears to be to resonate the structure with an inductance loaded with a ferrite or other permeable material; the ferrite is saturated by a control winding which changes its permeability, and thus, the frequency of the circuit, in accordance with the desired program.

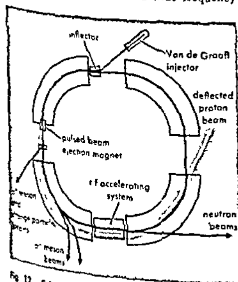


Fig. 12. Schematic diagram of the principal components of a proton synchrotron. Note the various possibilities of external neutral beams, charged beams, and extracted primary beams.

Figure 13 shows the proton synchrotron (called Bevatron) at the University of California Radiation Laboratory. The proton synchrotron at Brookhaven National Laboratory is called Cosmotron.

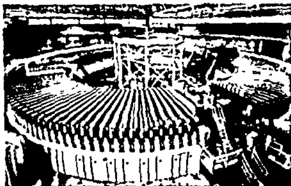


Fig. 13. The 6-Bev proton synchroton (Bevatron) of the University of California Radiation Laboratory, Berkeley. All four quadrants of the magnet are visible; the proton linear accelerator injector is shown at the lower right. (*University of California*)

Alternating-gradient synchrotron. As discussed previously, it is possible to achieve both radial and axial stability in a circular accelerator by rapidly alternating the sign of the radial gradient. Proton and electron accelerators based on this strong-focusing principle are called alternating-gradient synchrotrons (AGS). They differ from the conventional (weak-focusing) machines in several essential respects: (1) The radial and vertical particle excursions are considerably smaller because of the stronger focusing forces involved, this results in considerable reduction in magnet gap and, therefore, magnet size and cost. (2) Very severe mechanical tolerances must be met in order to avoid resonances, that is, integral relations between betatron and rotation frequencies. (3) Phase stability [illegible]

on the fact that the period of traversing an orbit increases with increase in particle momentum. Detailed analysis of AGS orbits indicates that this is in fact so only at low particle energies if high field gradients are used in the AGS; at higher energy, the period decreases with increasing particle momentum. This implies that the region of phase stability will shift from the part of the sine-wave voltage across the accelerating gap, decreasing in time, to a region increasing in time at a specific transition energy. At this energy, phase stability

field volume; hence, an accelerator of the synchrotron type requires a much smaller magnet than a machine requiring a magnet with a core extending to the center.

The variability of field required in the electron synchrotron is obtained by exciting the magnet from a source of alternating voltage. Since the energy stored in the magnetic field of the magnet is much larger than the energy dissipated as heat in the windings per accelerating cycle, economy requires that this energy be stored in a condenser bank between cycles. This can be achieved either by using the magnet and condenser bank as a parallel-tuned circuit across the power line or, if operation at a lower repetition rate is desired for reason of heat dissipation or economy, by switching the energy from the condenser bank to the magnet and back by high-power switching tubes such as *ignitrons*.

The synchrotron magnet must produce a field variable from tens of oersteds to more than 10,000 oersteds. The low-field part of the cycle gives difficulties on three scores: the effect of remanent fields, the effect of eddy-current fields, and the effect of static charges built up on the insulating walls of the "doughnut" which constitutes the vacuum chamber. Remanent field effects are minimized by careful randomization of the steel used in the poles and, occasionally, by circuit arrangements which demagnetize the iron between pulses. Eddy-current effects must be avoided by lamination of the iron (lamination sizes similar to those in conventional power transformers are used in most electron synchrotrons), and by manufacturing the vacuum chamber of insulating or very-low-conductivity materials (*see* EDDY CURRENT). Static charge effects are avoided by coating the inside of the chamber with a high-resistance (but conducting) material. In addition to these precautions, synchrotrons require a set of correcting coils which must be adjusted to correct the orbits at low magnetic field.

Since it is necessary to minimize the radial excursion of the beam, the synchrotron action must start at energies of 2 Mev or higher. Electron injection above this voltage can be done with an external electrostatic generator or electron linear accelerator and a pulsed electric inflector. If injection is made via an internal electron gun, the injection voltage is usually lower; in that case, the synchrotron's magnetic circuit is arranged so that the machine operates as a betatron until the necessary energy is reached.

The rf field to provide the necessary energy gain per turn and to make up the energy lost by the electrons because of their radiation in the circular orbit (the so-called synchrotron radiation) is usually provided by a gap in the rf cavity which is built into the doughnut chamber or installed in straight sections of the machine. The classical energy loss per turn due to radiation from an electron moving in a circle of radius r is given by

$$L = (4\pi/3)\,(r_0/r)\,\gamma^4 m_0 c^2 \qquad (18)$$

where $r_0 = 2.81 \times 10^{-13}$ cm is the classical electron radius, and the other symbols are the same as tho previously defined. This loss (varying as the cu of the energy if the magnetic field is held co stant) becomes the limiting factor to the ener attainable with the electron synchrotron (and mu sooner with the betatron): for a 5-Bev accelerato the energy loss is about 5 Mev/turn. In additio further complications arise at high energies b cause of the quantum nature of the emitted radi tion; these further increase the required accelera ing voltage. The radiation loss plus the require energy gain per turn defines the minimum acce erating voltage; in general, the minimum is e ceeded and thus the particle "rides" at a phase a gle defined by the magnetic field program and t energy of the rf system.

The output of electron synchrotrons is the x-ra beam produced by letting the electron strike a internal target. Schemes to extract the electro beam itself have been worked out, but not yet pu into practice (see Fig. 10).

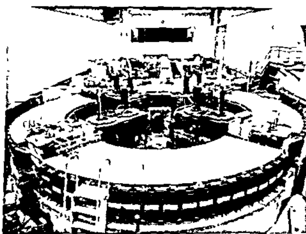


Fig. 10 The electron synchrotron of the California Institute of Technology Note the straight sections containing pumping manifolds, the accelerating system, and targets. This machine has reached 1.1-Bev energy. (*California Institute of Technology*)

Proton synchrotron. In the electron synchrotron, it is possible to accelerate highly relativistic ($\gamma >> 1$) electrons at nearly constant radius using a fixed accelerating frequency since the orbital velocity is nearly constant This is not possible for protons, since their velocity does not approach c for energies short of several billion electron volts (Bev). In order to realize the savings made possible by the use of a ring magnet—that is, in order to accelerate at constant orbit radius—Eqs. (16) and (17) show that both B (and hence ω_0) and ω_c must be changed together in order to keep r constant, and at the same time, increase the momentum. Figure 11 shows a plot of the required frequency (plotted as $\omega_c r/c$) as a function of the magnetic field (plotted as $\omega_0 r/c$) required to keep the radius constant. This is the principle of the proton synchrotron. In other respects, operation is identical to that of the electron synchrotron with the exception that radiation losses are totally negligible.

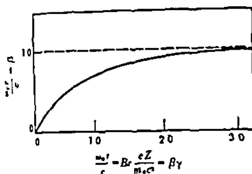


Fig 11 Plot of particle velocity β (in units of the velocity of light c) against particle momentum $\beta\gamma$ (in units of rest energy divided by c) The particle velocity is a measure of the required frequency ω_c of the accelerating system, and the momentum is a measure of the magnetic field The curve thus represents the required program of magnetic field and frequency

There are, however, many additional practical differences Proton synchrotrons usually operate at a lower rate of rise of magnetic field, and hence, the magnet lamination can be thicker without excessive eddy-current effects. Also, the vacuum chamber can be fabricated of thin stainless steel or other low conductivity metal. Injection is accomplished by a separate electrostatic accelerator or proton linear accelerator and a pulsed electric inflection system Proton synchrotrons have been constructed in the form of four quadrants separated by straight sections The inflection electrodes, the rf accelerating cavities, and other apparatus are located in the field free straight sections Figure 12 shows a typical arrangement.

The rf system and its control present special problems since the accelerating cavity (or electrode) must be modulated over a wide frequency range in accordance with the program shown in Fig 11. To give the accelerating structure enough bandwidth to handle the frequency range would require excessive power. The best solution appears to be to resonate the structure with an inductance loaded with a ferrite or other permeable material; the ferrite is saturated by a control winding which changes its permeability, and thus, the frequency of the circuit, in accordance with the desired program.

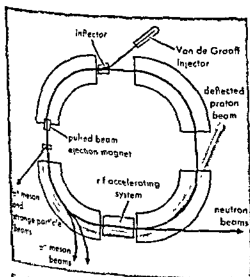


Fig 12 Schematic diagram of the principal components of a proton synchrotron Note the various possibilities of external neutral beams, charged beams, and extracted primary beams

Figure 13 shows the proton synchrotron (called Bevatron) at the University of California Radiation Laboratory. The proton synchrotron at Brookhaven National Laboratory is called Cosmotron.

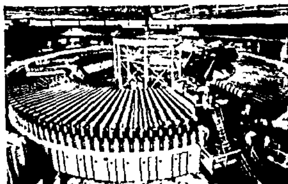


Fig 13 The 6-Bev proton synchrotron (Bevatron) of the University of California Radiation Laboratory, Berkeley All four quadrants of the magnet are visible; the proton linear accelerator injector is shown at the lower right (University of California)

Alternating-gradient synchrotron [illegible]

and electron accelerators based on this strong focusing principle are called alternating-gradient synchrotrons (AGS) They differ from the conventional (weak-focusing) machines in several essential respects (1) The radial and vertical particle excursions are considerably smaller because of the stronger focusing forces involved, this results in considerable reduction in magnet gap and, therefore, magnet size and cost (2) Very severe mechanical tolerances must be met in order to avoid resonances, that is, integral relations between betatron and rotation frequencies. (3) Phase stability considerations [illegible]

on the fact that the period of traversing an orbit increases with increase in particle momentum Detailed analysis of AGS orbits indicates that this is in fact so only at low particle energies if high field gradients are used in the AGS; at higher energy, the period decreases with increasing particle momentum This implies that the region of phase stability will shift from the part of the sine-wave voltage across the accelerating gap, decreasing in time, to a region increasing in time at a *specific* transition energy. At this energy, phase stability

field volume; hence, an accelerator of the synchrotron type requires a much smaller magnet than a machine requiring a magnet with a core extending to the center.

The variability of field required in the electron synchrotron is obtained by exciting the magnet from a source of alternating voltage. Since the energy stored in the magnetic field of the magnet is much larger than the energy dissipated as heat in the windings per accelerating cycle, economy requires that this energy be stored in a condenser bank between cycles. This can be achieved either by using the magnet and condenser bank as a parallel-tuned circuit across the power line or, if operation at a lower repetition rate is desired for reason of heat dissipation or economy, by switching the energy from the condenser bank to the magnet and back by high-power switching tubes such as ignitrons.

The synchrotron magnet must produce a field variable from tens of oersteds to more than 10,000 oersteds. The low-field part of the cycle gives difficulties on three scores: the effect of remanent fields, the effect of eddy-current fields, and the effect of static charges built up on the insulating walls of the "doughnut" which constitutes the vacuum chamber. Remanent field effects are minimized by careful randomization of the steel used in the poles and, occasionally, by circuit arrangements which demagnetize the iron between pulses. Eddy-current effects must be avoided by lamination of the iron (lamination sizes similar to those in conventional power transformers are used in most electron synchrotrons), and by manufacturing the vacuum chamber of insulating or very-low-conductivity materials (*see* EDDY CURRENT). Static charge effects are avoided by coating the inside of the chamber with a high-resistance (but conducting) material. In addition to these precautions, synchrotrons require a set of correcting coils which must be adjusted to correct the orbits at low magnetic field

Since it is necessary to minimize the radial excursion of the beam, the synchrotron action must start at energies of 2 Mev or higher. Electron injection above this voltage can be done with an external electrostatic generator or electron linear accelerator and a pulsed electric inflector If injection is made via an internal electron gun, the injection voltage is usually lower; in that case, the synchrotron's magnetic circuit is arranged so that the machine operates as a betatron until the necessary energy is reached

The rf field to provide the necessary energy gain per turn and to make up the energy lost by the electrons because of their radiation in the circular orbit (the so-called synchrotron radiation) is usually provided by a gap in the rf cavity which is built into the doughnut chamber or installed in straight sections of the machine. The classical energy loss per turn due to radiation from an electron moving in a circle of radius r is given by

$$L = (4\pi/3)(r_0/r)\gamma^4 m_0 c^2 \qquad (18)$$

where $r_0 = 2.81 \times 10^{-13}$ cm is the classical electron radius, and the other symbols are the same as those previously defined. This loss (varying as the cube of the energy if the magnetic field is held constant) becomes the limiting factor to the energy attainable with the electron synchrotron (and much sooner with the betatron): for a 5-Bev accelerator the energy loss is about 5 Mev/turn. In addition further complications arise at high energies because of the quantum nature of the emitted radiation; these further increase the required accelerating voltage. The radiation loss plus the required energy gain per turn defines the minimum accelerating voltage; in general, the minimum is exceeded and thus the particle "rides" at a phase angle defined by the magnetic field program and the energy of the rf system.

The output of electron synchrotrons is the x-ray beam produced by letting the electron strike an internal target. Schemes to extract the electron beam itself have been worked out, but not yet put into practice (see Fig. 10).

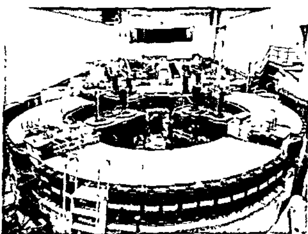


Fig. 10 The electron synchrotron of the California Institute of Technology. Note the straight sections containing pumping manifolds, the accelerating system, and targets. This machine has reached 1.1-Bev energy. (California Institute of Technology)

Proton synchrotron. In the electron synchrotron, it is possible to accelerate highly relativistic ($\gamma >> 1$) electrons at nearly constant radius using a fixed accelerating frequency since the orbital velocity is nearly constant. This is not possible for protons, since their velocity does not approach c for energies short of several billion electron volts (Bev). In order to realize the savings made possible by the use of a ring magnet—that is, in order to accelerate at constant orbit radius—Eqs. (16) and (17) show that both B (and hence ω_0) and ω_c must be changed together in order to keep r constant, and at the same time, increase the momentum. Figure 11 shows a plot of the required frequency (plotted as $\omega_c r/c$) as a function of the magnetic field (plotted as $\omega_0 r/c$) required to keep the radius constant. This is the principle of the proton synchrotron. In other respects, operation is identical to that of the electron synchrotron with the exception that radiation losses are totally negligible.

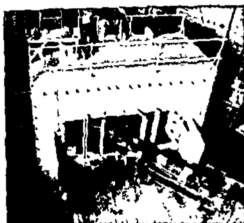


Fig. 16 The 300-Mev betatron of the University of Illinois. It is the highest energy betatron built as of 1959. The big external iron yoke (light-colored) is the magnetic return for the flux core. The guide-field structure is the smaller dark-colored circular structure. (University of Illinois)

magnetic shield which permits the electrons in specific orbits to escape from the field.

Equation (19) does not allow for radiation loss; hence, the practical limit of the betatron is set when it becomes impossible to program $\dot{\phi}$ to a rate sufficient to compensate for the radiation loss. This limit is near 500 Mev (see Fig. 16).

Fixed field accelerators. With the exception of the ordinary *cyclotron*, *all* the circular (magnetic) accelerators described employ time-varying fields with the consequent need of laminated construction and large external equipment for storage of the magnetic [illegible]
cyclotro [illegible]
stability [illegible]
tain syn [illegible]
overcome these problems have gone in two directions: (1) to design a field in which cyclotron resonance at a fixed orbital frequency and radial and axial focusing can be maintained simultaneously at all radii, and (2) to design fixed field machines using alternating-gradient focusing with phase-stable operation and variable frequency where the frequency program controls the radial position of the orbit.

The first method has been realized in the Thomas cyclotron which operates like an ordinary cyclotron, but employs a magnetic field variable in azimuth in such a way that stability is obtained. This machine offers the possibility of stable operation to near 1 Bev at very high intensity with continuous beam, but is quite expensive to construct. Another machine in this class is the so-called microtron for electrons, which has been constructed to several-Mev energy. The microtron employs a magnetic field such that the time of successive revolutions of the particles increases by exactly 1 *cycle* of the accelerating rf voltage; synchronism is thus maintained.

The second method (*b*) is the basis of the so-called fixed field, alternating-gradient (FFAG) machines. These are still (1960) in the development stages; several models have been operated successfully.

LINEAR ACCELERATORS

Principles of operation. Linear accelerators accelerate particles in a straight line by means of suitable rf fields. Since a given field location is traversed only once, such fields must be produced along the entire particle orbit; hence, very high field strengths are necessary, and this implies very high rf power levels.

Possible synchronism between particles and field is attained by either of two methods: traveling-wave acceleration, wherein a wave with an accelerating field component is produced whose phase velocity is equal to the particle velocity; or standing-wave acceleration, wherein a standing-wave pattern is produced which would alternately accelerate and decelerate the particles. [illegible]

[illegible]

ticle can be accelerated in a structure of length L fed by rf power sources operating at a wavelength λ and feeding in a power P is given by

$$T = K\,(PL)^{1/2}\lambda^{-1/4} \tag{20}$$

where K is a constant depending only slightly on the structure chosen and also on whether a standing-wave or traveling-wave structure is used; hence length and power are equally instrumental in attaining high energies. Short wavelengths are advantageous from the point of view of power economy, but factors opposing this view are (1) in principle, higher power levels are available from microwave tubes at longer wavelengths and higher powers can be dissipated in the accelerator; (2) the tolerances of the mechanical structure are easier to meet at longer wavelengths; (3) higher currents can be accelerated at longer wavelengths. In practice, the wavelengths used are 3–30 cm for electron accelerators, 100–200 cm for proton accelerators, and 500–1000 cm for heavy-ion accelerators. The high power levels involved require that linear accelerators be operated as pulsed rather than as continuously operated machines; note that this requirement is imposed by practical and eco-

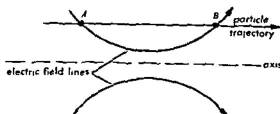


Fig. 17. Phase focusing in the linear accelerator in the absence of grids. If the field is increasing during particle transit across an accelerating gap, the defocusing momentum imparted at B will be greater than the focusing momentum at A. The action is thus defocusing but is phase-stable.

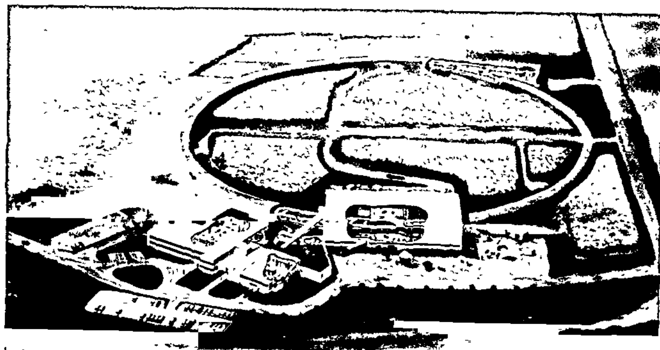


Fig. 14 Artist's drawing of the 25-Bev alternating-gradient proton synchrotron under construction at Brookhaven National Laboratory. The linear accelerator injector is at the upper right of the picture; the indoor and outdoor experimental areas are at the front.

will disappear briefly. To recapture the particles, a shift in phase of the rf voltage is required Successful passage of particles through this "phase transition" region has been achieved in an electron model and in the alternating-gradient synchrotron of the European Council for Nuclear Research (CERN), Geneva, which is now operating at 29 Bev. A similar machine was scheduled for completion in 1960 at the Brookhaven National Laboratory (Fig. 14).

Betatron. The betatron (or magnetic-induction accelerator) is a circular electron accelerator in which a magnet with radial gradient provides the guide-field, but in which the energy is given to the particles through a time-varying magnetic flux linking the orbits The action can be considered to be that of a transformer in which the primary winding is composed of ordinary wire turns while the secondary is the electron beam.

A certain condition must be met if the particle orbit radius r is to be constant while the flux ϕ and the magnetic induction B at the orbit are changing at rates $\dot{\phi}$ and $\dot{B}$. The energy gain per turn is given by $e\dot{\phi}$, according to Faraday's law. Hence,

$$e\dot{\phi} = (mc^2\dot{\gamma})(2\pi r/v) = 2\pi r(dp/dt) = 2\pi r^2 e\dot{B} \quad (19)$$

by differentiating the appropriate relativistic equations. Equation (19) shows that the flux through the orbit must change at twice the rate as the product of orbit area times the flux density at the orbit. This is known as the betatron condition. Note that Eq (19), which determines the radius at which the particles circulate, refers only to the time rates of the field and the flux; economy can be realized by letting the flux range from negative to positive values while the guide-field excursion is unilateral. This is called flux-biased operation. Figure 15 is a schematic diagram of the betatron with a separate guide-field and flux drive.

Electron injection into a betatron is usually by an electron gun at the edge of the vacuum chamber The current obtainable depends critically on the ability of the electrons to avoid hitting the injector on the initial turns until the increasing field contracts the orbit. The output beam is usually the x-ray beam produced when the electrons strike a target; however, ejection of an external electron beam has been accomplished for the lower-energy betatrons by a "peeler," which is a carefully shaped

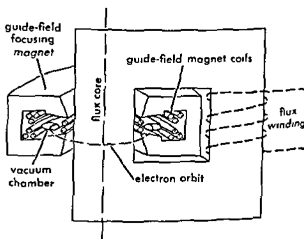


Fig 15. Schematic diagram of the basic components of a betatron The energy is given to the particle through the time-varying flux linking the orbit produced in the flux core. The orbit is bent and focused by the guide-field The orbit field and the central flux are programmed together to obey the betatron condition of maintaining a constant orbit radius.

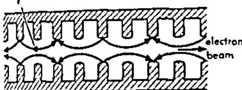


Fig 20 Electric field configuration in a traveling-wave electron linear accelerator. The wave guide shown is loaded by four disks per wavelength (other spacings are possible) to obtain a phase velocity to match the speed of the particle

ing dominant in electrostatic accelerators, but very small here, is neglected This incompatability of focusing and phase stability appears to be a serious obstacle to the operation of a linear accelerator, but it can be circumvented in various designs as follows: (1) A very short accelerator can be operated even though unstable (2) Charge can be included in the beam to terminate the lines in Fig. 17 inside the beam, thus producing a convergent action even at phase-stable transit phases; this charge can be induced on grids placed in the field. (3) As the particle velocity approaches c, the apparent incompatibility becomes irrelevant because, for relativistic velocities, the action of the radial electric time-varying field is almost canceled by the accompanying time-varying magnetic field, and also, because the velocity is almost invariable, so that the particles are in neutral equilibrium both longitudinally and transversely. (4) External magnetic (either solenoidal or strong-focusing) or electrostatic lenses can be used. *See* MAGNETIC LENS.

Proton and ion accelerators. Existing proton and heavy ion linear accelerators are of the standing-wave type. The wave pattern is set up in a cavity with an electric field configuration as shown in Fig. 18 The shaded areas around the beam represent metallic drift tubes; no field is present inside the tubes. The distance between tube center lines is $v\lambda/c$, where λ is the exciting wavelength and v the mean particle velocity between the center lines. If this condition is fulfilled, the particles will cross each gap at the same phase, and phase-stable acceleration will result. Focusing is obtained by grids at the entrance of each tube or, in all recent designs, by lenses built into the tubes. The cavity and drift tubes must be constructed to make the structure resonant for the mode shown (*see* CAVITY RESONATOR). The resultant cavity is excited by external oscillators or amplifiers Several cavities of the type shown can be coupled together and excited in synchronism Ions are injected into the accelerator by an electrostatic machine; the beam emerges at the output end, and can be directed to the experimental targets with suitable magnetic arrangements (see Fig. 19).

Electron accelerators. Almost all electron linear accelerators are of the traveling-wave type. The wave is produced in a loaded wave guide, which is a wave guide excited in the TM mode (having a longitudinal electric field component) and loaded by disks or other means to produce the correct phase velocity to match the particle An unloaded or uniform wave guide is not suitable, since the phase velocity in such a guide exceeds the velocity of light. *See* PHASE VELOCITY, WAVE GUIDE.

A typical field configuration is shown in Fig. 20; the field pattern should be visualized as translating along the axis with a velocity equal to that of the electron shown The electron velocities approach c very rapidly, hence, except for the first few feet of the machine, all radial forces can be neglected. This means that the momentum component trans-

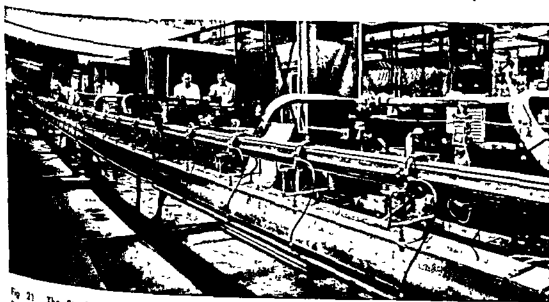


Fig 21 The Stanford University Mark III electron linear accelerator with shielding removed Note the wave-guide feeds which transmit the microwave power This accelerator has reached 700 Mev (Stanford University)

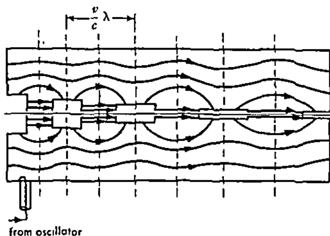


Fig 18. Electric field configuration in the standing-wave proton accelerator employing a cavity resonant in the TM_{01} mode. Drift tubes are shown placed to shield the particles when the phase is decelerating The length of the drift tubes is designed such that the particles cross each gap at approximately the same phase

nomic considerations rather than by fundament principles of operation, as in the case of the phas stable circular machines.

Orbit dynamics. The focusing action and phas stability of linear accelerators in the absence external lenses that could be placed around th machine will now be discussed. Consider a regio in space where the accelerating field lines increas in time as the particle crosses the region (Fig. 17) assume that no lines end inside the beam. It is clea that there will be a net defocusing action since in an increasing field, the transverse momentum im parted at B is greater than that imparted at A However, the action is phase-stable, since a parti cle arriving late will receive a larger acceleratio and will catch up on the next cycle. If the particl *crosses during a decreasing part, the conclusion* are reversed. Thus, focusing and phase stabilit in a linear accelerator are incompatible. This ca be proved by formal mathematical methods; th only approximation is that the second-order focus

Fig. 19. Inside view of the cavity of the heavy-ion linear accelerator at the University of California Radiation Laboratory, Berkeley. Note the drift-tube structure and the rf coupling loops on the edge of the tank (*University of California*)

highest speed The signal may be 10^{-9} sec long and the dead time zero Geiger counters are not so fast, with a dead time of about 10^{-4} sec Most cloud chambers have a dead time of more than 1 min. Other particle detectors have speeds intermediate between these extremes.

Proportionality. When the magnitude of the electrical signal is proportional to the energy loss by the particle in the detector, the counter is said to be proportional in its characteristics Scintillation counters are proportional, but Geiger-Müller counters are not.

Measurements possible. Cloud chambers and bubble chambers, although inconvenient to use, can give much more detailed information than other detectors about the particles which pass through them. If a chamber is placed in a magnetic field, the momentum of the particle may be determined from the curvature of the track. Measurements of velocity may also be obtained from cloud-chamber and bubble-chamber tracks. *See* LOW-LEVEL COUNTING [W.B.F.]

Bibliography: E. Segré (ed), *Experimental Nuclear Physics*, vol. 1, 1953.

Particle properties

Particles [illegible] A powder is a bulk assemblage of fine [illegible] to a dilute suspension of fine particles. Correspondingly, a hydrosol is a suspension of particles in a liquid

Particle size. The characteristics of particles are primarily related to their size, although other factors, such as shape, density, and composition, may also be important. Size is generally expressed in terms of some representative, average, or effective dimension of the particle The most widely used unit of particle size is the micron, defined as 1/1000 mm (1/25,400 in.) and given the symbol μ Another common method is to designate the screen mesh that has an aperture corresponding to the particle size The screen mesh usually refers to the number of screen openings per unit length or area, several screen standards are in general use, the two most common in the United States being the U.S. Standard and the Tyler Standard screen scales.

Natural and synthetic particles rarely consist of particles of only one size; a range of particle size is usually encountered. Thus, to describe a particulate material, it is necessary to specify not only the size and shape of individual particles but also the relative amounts of each size. This is known as a size distribution, obtained by means of a particle size analysis. A wide variety of particle size analysis methods is available. It should be noted that most methods of size analysis do not measure particle size Instead, they measure some other properties of the particles which are then converted to an apparent or effective size by means of analytical or empirical relationships

Mechanical dispersoids are formed by comminution, decrepitation, or disintegration of larger masses of material, as by grinding of solids or spraying of liquids, and usually involve a wide particle-size distribution. Condensed dispersoids are formed by condensation of the vapor phase or as the product of a vapor-phase reaction and are usually very fine, and often relatively uniform in size. Condensed dispersoids and fine mechanical dispersoids generally tend to flocculate or agglomerate to form loose clusters of larger particle size

Particle concentration. The concentration of particles in a dispersoid is often expressed in terms of grains per cubic foot of gas, where 7000 grains = 1 lb. In air pollution work, concentrations are sometimes expressed as grams per cubic meter (1 [illegible] dust concentr[illegible] 01 to 100 grain/ft³, with 1–10 grain/ft³ being common in process ventilation work. In air-conditioning applications, concentrations of particulate matter will generally range from 0.1 to 10 grains/1000 ft³. In operations such as pneumatic conveying, the concentrations will range from 0 1 to 50 lb solid/lb gas (about 50–20,000 grain/ft³) In fluidized-solid systems, concentrations will frequently exceed 1000 lb solid/lb gas.

Figure 1 shows a listing of particle sizes of common materials and related items, as well as methods of size analysis. Figure 2 shows a summary of concentrations of particulates suspended in air.

Particle mechanics. A particle moving in a fluid encounters a frictional resistance

$$F_r = C_D A_p \rho \frac{u_r^2}{2}$$

where F_r is the resisting force in dynes, C_D is the dimensionless drag coefficient; A_p is the area of the particle projected in a plane normal to the direction of motion in cm², ρ is the fluid density in g/cm³, and u_r is the velocity of the particle relative to the fluid in cm/sec. The drag coefficient C_D is a function of the particle shape, the proximity of bounding surfaces, and the Reynolds number N_{Re}($= D_p u_r \rho / \mu$, where D_p is the particle diameter or other representative dimension in cm; and μ is the fluid viscosity in poises). Drag coefficient data are available for a wide variety of particle shapes

If a particle suspended in a fluid is acted upon by a force, it will accelerate to a terminal velocity, at which the resisting force due to fluid friction just balances the applied force If a particle falls under the action of gravity, this velocity is known as the terminal gravitational settling velocity

$$u_t = \sqrt{2 g_L M_p (\rho_p - \rho) / \rho \rho_p A_p C_D}$$

where u_t is the settling velocity in cm/sec; g_L is the acceleration due to gravity, 980 7 cm²/sec, M_p is the particle mass in grams; ρ_p is particle density in g/cm³; and other terms are as previously defined

verse to the axis remains constant, and since the longitudinal component increases continuously, the angle of divergence of the beam decreases continuously; for uniform energy gain per unit length, this corresponds to a beam angle varying as the inverse of the distance along the machine and to a beam radius increasing logarithmically. A set of very small external magnetic lenses can change this beam diameter increase to a decrease.

The accelerator wave guide must be constructed to very close tolerances to control the phase velocity to the required accuracy, since there is no phase stability once the energy has become several times the rest energy of the electron.

Particles are injected from an electron gun which can easily inject electrons at a velocity of $\approx c/2$. The first accelerator section can be a special "bunching" device to concentrate the particles near the crest of the traveling wave of the succeeding sections.

The energy possibilities of the electron linear accelerator are not limited by fundamental considerations, since electrons accelerated in a straight line do not lose energy by radiation by an appreciable amount. Since the operating frequency is in the microwave region, the possible performance of the accelerators is closely tied to the available power sources. Some of the earlier machines were powered by magnetron oscillators; present machines are powered by klystron amplifiers driven from a common master oscillator (see Fig. 21). *See* MICROWAVE TUBE. [W.K.H.P.]

Bibliography: G. A. Behrman, *Particle Accelerators*, Univ. Calif. (Berkeley) Radiation Lab. Rept. UCRL-8050, 1958; D. E. Gray (ed.), *American Institute of Physics Handbook*, 1957; M. S. Livingston, *High Energy Accelerators*, 1954; *Proceedings of International Conference on High-Energy Accelerators and Instrumentation (CERN), Geneva, 1959*; E. Segrè (ed.), *Experimental Nuclear Physics*, vol. 3, 1959.

Particle detector

A device used to indicate the presence of fast-moving charged atomic or nuclear particles. When such a particle passes through a particle detector, it creates an electrical disturbance in the detector which can be observed and recorded. Particle detectors are used in research in nuclear and atomic physics and to detect cosmic rays. They are important in exploration for radioactive minerals, since they detect the particles emitted from radioactive substances.

Kinds of detectors. Many different types of particle detectors are available. For detailed information, *see* BUBBLE CHAMBER; CERENKOV RADIATION; CLOUD CHAMBER; CRYSTAL COUNTER; GEIGER-MÜLLER COUNTER; IONIZATION CHAMBER; SCINTILLATION COUNTER; SCINTILLATION DETECTOR, LIQUID; SPARK COUNTER. For information on specific methods of detection, *see* BETA RAYS, GAMMA RAYS; NEUTRON.

Another type of particle detector is the photographic emulsion. Fast charged particles passing through silver bromide grains in photographic emulsion leave the grains developable. Radioactive materials absorbed in tissue or bone reveal their distribution if the substance is placed in contact with a photographic plate, giving an autoradiograph; for details, *see* AUTORADIOGRAPHY. High-energy charged particles leave trails of developable grains which may be followed through the emulsion if they are observed in a high-power microscope. Nuclear emulsion plates are important particle detectors for research in high-energy nuclear physics. *See* PHOTOGRAPHY.

The particles detected may be electrons, protons, α-particles (or other atomic nuclei), and various types of mesons. They must be moving fast enough to penetrate the walls of the detector and still have enough energy to produce the electrical disturbance. The particles must be charged; neutral particles produce no electrical disturbance directly. Neutrons, photons, and other neutral particles may only be detected if they create or set free charged particles which can then be observed with a particle detector.

Particle detectors generally consist of a sensitive element such as a Geiger-Müller (or Geiger) counter or a scintillator together with amplifiers to increase the intensity of the signal created when a particle passes through so that it will operate a mechanical register or an electrical meter. When cloud chambers or bubble chambers are used as particle detectors, the tracks of fast charged particles passing through the chamber are made visible and are photographed.

Information provided. In addition to providing the basic information concerning the presence of a fast-moving charged particle, particle detectors may provide much more detailed information. Depending on the detector used, information may be obtained on:

1 Number traversing detector per second (counting rate)
2 Precise time when particle traversed detector
3 Precise position of particle
4 Energy lost by particle
5 In case of radioactive particle, lifetime of particle
6 Momentum of particle
7 Velocity of particle

Often different particle detectors are used together to obtain information. For example, Geiger-Müller counters are used in conjunction with cloud chambers to provide counter-controlled photographs of particle tracks

Detector characteristics. The characteristics of particle detectors which are important in determining the type of detector to be used in a given application are speed, proportionality, and the amount of data desired

Speed. The maximum counting rate of a particle detector is determined by the length of the electrical signal created and the time required for the detector to recover from a given impulse (the so-called dead time). Scintillation counters have the

highest speed The signal may be 10^{-9} sec long and the dead time zero Geiger counters are not so fast, with a dead time of about 10^{-4} sec Most cloud chambers have a dead time of more than 1 min. Other particle detectors have speeds intermediate between these extremes

Proportionality. When the magnitude of the electrical signal is proportional to the energy loss by the particle in the detector, the counter is said to be proportional in its characteristics Scintillation counters are proportional, but Geiger-Müller counters are not.

Measurements possible. Cloud chambers and bubble chambers, although inconvenient to use, can give much more detailed information than other detectors about the particles which pass through them If a chamber is placed in a magnetic field, the momentum of the particle may be determined from the curvature of the track. Measurements of velocity may also be obtained from cloud-chamber and bubble-chamber tracks. *See* LOW-LEVEL COUNTING [W.B F]

Bibliography: E. Segré (ed), *Experimental Nuclear Physics*, vol 1, 1953

Particle properties

Particles are solids or liquids in a subdivided state. Bec...

... particles in a fluid An aerosol is a suspension of particles, either liquid or solid, in a gas; the term usually refers to a dilute suspension of fine particles Correspondingly, a hydrosol is a suspension of particles in a liquid.

Particle size. The characteristics of particles are primarily related to their size, although other factors, such as shape, density, and composition, may also be important Size is generally expressed in terms of some representative, average, or effective dimension of the particle The most widely used unit of particle size is the micron, defined as 1/1000 mm (1/25,400 in) and given the symbol μ Another common method is to designate the screen mesh that has an aperture corresponding to the particle size The screen mesh usually refers to the number of screen openings per unit length or area; several screen standards are in general use, the two most common in the United States being the U S Standard and the Tyler Standard screen scales.

Natural and synthetic particles rarely consist of particles of only one size; a range of particle size is usually encountered Thus, to describe a particulate material, it is necessary to specify not only the size and shape of individual particles but also the relative amounts of each size. This is known as a size distribution, obtained by means of a particle size analysis A wide variety of particle size analysis methods is available It should be noted that most methods of size analysis do not measure particle size. Instead, they measure some other properties of the particles which are then converted to an apparent or effective size by means of analytical or empirical relationships.

Mechanical dispersoids are formed by comminution, decrepitation, or disintegration of larger masses of material, as by grinding of solids or spraying of liquids, and usually involve a wide particle-size distribution. Condensed dispersoids are formed by condensation of the vapor phase or as the product of a vapor-phase reaction and are usually very fine, and often relatively uniform in size Condensed dispersoids and fine mechanical dispersoids generally tend to flocculate or agglomerate to form loose clusters of larger particle size.

Particle concentration. The concentration of particles in a dispersoid is often expressed in terms of grains per cubic foot of gas, where 7000 grains = 1 lb. In air pollution work, concentrations are sometimes expressed as grams per cubic meter (1 g/m³ = 0.436 grains/ft³). Process dust concentrations normally range from 0 01 to 100 grain/ft³, with 1–10 grain/ft³ being common in process ventilation work. In air-conditioning applications, concentrations of particulate matter will generally range from 0 1 to 10 grains/1000 ft³. In operations such as pneumatic conveying, the concentrations will range ... gas ... syste... 1000 lb solid/lb gas

Figure 1 shows ...

... centrations of particulates suspended in air.

Particle mechanics. A particle moving in a fluid encounters a frictional resistance

$$F_r = C_D A_p \rho \frac{u_r^2}{2}$$

where F_r is the resisting force in dynes, C_D is the dimensionless drag coefficient; A_p is the area of the particle projected in a plane normal to the direction of motion in cm²; ρ is the fluid density in g/cm³, and u_r is the velocity of the particle relative to the fluid in cm/sec The drag coefficient C_D is a function of the particle shape, the proximity of bounding surfaces, and the Reynolds number N_{Re} (= $D_p u_r \rho/\mu$, where D_p is the particle diameter or other ...

... balances the applied force If a particle falls under the action of gravity, this velocity is known as the terminal gravitational settling velocity

$$u_t = \sqrt{2g_L M_p(\rho_p - \rho)/\rho\rho_p A_p C_D}$$

where u_t is the settling velocity in cm/sec, g_L is the acceleration due to gravity, 980 7 cm²/sec, M_p is the particle mass in grams; ρ_p is particle density in g/cm³, and other terms are as previously defined.

verse to the axis remains constant, and since the longitudinal component increases continuously, the angle of divergence of the beam decreases continuously; for uniform energy gain per unit length, this corresponds to a beam angle varying as the inverse of the distance along the machine and to a beam radius increasing logarithmically. A set of very small external magnetic lenses can change this beam diameter increase to a decrease.

The accelerator wave guide must be constructed to very close tolerances to control the phase velocity to the required accuracy, since there is no phase stability once the energy has become several times the rest energy of the electron.

Particles are injected from an electron gun which can easily inject electrons at a velocity of $\approx c/2$. The first accelerator section can be a special "bunching" device to concentrate the particles near the crest of the traveling wave of the succeeding sections.

The energy possibilities of the electron linear accelerator are not limited by fundamental considerations, since electrons accelerated in a straight line do not lose energy by radiation by an appreciable amount. Since the operating frequency is in the microwave region, the possible performance of the accelerators is closely tied to the available power sources. Some of the earlier machines were powered by magnetron oscillators; present machines are powered by klystron amplifiers driven from a common master oscillator (see Fig. 21). *See* MICROWAVE TUBE. [W.K.H.P.]

Bibliography: G A. Behrman, *Particle Accelerators*, *Univ. Calif. (Berkeley) Radiation Lab. Rept.* UCRL-8050, 1958; D. E. Gray (ed.), *American Institute of Physics Handbook*, 1957; M. S. Livingston, *High Energy Accelerators*, 1954; *Proceedings of International Conference on High-Energy Accelerators and Instrumentation (CERN), Geneva*, 1959; E. Segrè (ed.), *Experimental Nuclear Physics*, vol. 3, 1959.

Particle detector

A device used to indicate the presence of fast-moving charged atomic or nuclear particles. When such a particle passes through a particle detector, it creates an electrical disturbance in the detector which can be observed and recorded. Particle detectors are used in research in nuclear and atomic physics and to detect cosmic rays. They are important in exploration for radioactive minerals, since they detect the particles emitted from radioactive substances.

Kinds of detectors. Many different types of particle detectors are available. For detailed information, *see* BUBBLE CHAMBER; CERENKOV RADIATION; CLOUD CHAMBER; CRYSTAL COUNTER; GEIGER-MÜLLER COUNTER; IONIZATION CHAMBER; SCINTILLATION COUNTER; SCINTILLATION DETECTOR, LIQUID; SPARK COUNTER. For information on specific methods of detection, *see* BETA RAYS; GAMMA RAYS; NEUTRON.

Another type of particle detector is the photographic emulsion. Fast charged particles passing through silver bromide grains in photographic emulsion leave the grains developable. Radioactive materials absorbed in tissue or bone reveal their distribution if the substance is placed in contact with a photographic plate, giving an autoradiograph; for details, *see* AUTORADIOGRAPHY. High-energy charged particles leave trails of developable grains which may be followed through the emulsion if they are observed in a high-power microscope. Nuclear emulsion plates are important particle detectors for research in high-energy nuclear physics. *See* PHOTOGRAPHY.

The particles detected may be electrons, protons, α-particles (or other atomic nuclei), and various types of mesons. They must be moving fast enough to penetrate the walls of the detector and still have enough energy to produce the electrical disturbance. The particles must be charged; neutral particles produce no electrical disturbance directly. Neutrons, photons, and other neutral particles may only be detected if they create or set free charged particles which can then be observed with a particle detector.

Particle detectors generally consist of a sensitive element such as a Geiger-Müller (or Geiger) counter or a scintillator together with amplifiers to increase the intensity of the signal created when a particle passes through so that it will operate a mechanical register or an electrical meter. When cloud chambers or bubble chambers are used as particle detectors, the tracks of fast charged particles passing through the chamber are made visible and are photographed

Information provided. In addition to providing the basic information concerning the presence of a fast-moving charged particle, particle detectors may provide much more detailed information. Depending on the detector used, information may be obtained on:

1. Number traversing detector per second (counting rate)
2. Precise time when particle traversed detector
3. Precise position of particle
4. Energy lost by particle
5. In case of radioactive particle, lifetime of particle
6. Momentum of particle
7. Velocity of particle

Often different particle detectors are used together to obtain information. For example, Geiger-Müller counters are used in conjunction with cloud chambers to provide counter-controlled photographs of particle tracks.

Detector characteristics. The characteristics of particle detectors which are important in determining the type of detector to be used in a given application are speed, proportionality, and the amount of data desired

Speed. The maximum counting rate of a particle detector is determined by the length of the electrical signal created and the time required for the detector to recover from a given impulse (the so-called dead time). Scintillation counters have the

chambers have a dead time of more than 1 min. Other particle detectors have speeds intermediate between these extremes.

Proportionality When the magnitude of the electrical signal is proportional to the energy loss by the particle in the detector, the counter is said to be proportional in its characteristics Scintillation counters are proportional, but Geiger-Müller counters are not.

Measurements possible. Cloud chambers and bubble chambers, although inconvenient to use, can give much more detailed information than other detectors about the particles which pass through them If a chamber is placed in a magnetic field, the momentum of the particle may be determined from the curvature of the track. Measurements of velocity may also be obtained from cloud-chamber and bubble-chamber tracks. *See* LOW-LEVEL COUNTING [W B F.]

Bibliography· E. Segré (ed.), *Experimental Nuclear Physics*, vol 1, 1953

Particle properties

Particles are solids or liquids in a subdivided state. Because of this subdivision, particles exhibit many special characteristics which are negligible in the bulk material A powder is a bulk assemblage of fine, dry, solid particles A dispersoid is a suspen-

[illegible]

fine particles Correspondingly, a hydrosol is a suspension of particles in a liquid.

Particle size. The characteristics of particles are primarily related to their size, although other factors, such as shape, density, and composition, may also be important. Size is generally expressed in terms of some representative, average, or effective dimension of the particle. The most widely used unit of particle size is the micron, defined as 1/1000 mm (1/25,400 in.) and given the symbol μ Another common method is to designate the screen mesh that has an aperture corresponding to the particle size. The screen mesh usually refers to the number of screen openings per unit length or area; several screen standards are in general use, the two most common in the United States being the U.S. Standard and the Tyler Standard screen scales.

Natural and synthetic particles rarely consist of particles of only one size; a range of particle size is usually encountered. Thus, to describe a particulate material, it is necessary to specify not only the size and shape of individual particles but also the relative amounts of each size. This is known as a size distribution, obtained by means of a particle size analysis. A wide variety of particle size analysis methods is available. It should be noted that [illegible] methods of size analysis do not measure particle size. Instead, [illegible] other properties [illegible] converted to an apparent or effective size by means of analytical or empirical relationships.

Mechanical dispersoids are formed by comminution, decrepitation, or disintegration of larger masses of material, as by grinding of solids or spraying of liquids, and usually involve a wide particle-size distribution. Condensed dispersoids are formed by condensation of the vapor phase or as the product of a vapor-phase reaction and are usually very fine, and often relatively uniform in size. Condensed dispersoids and fine mechanical dispersoids generally tend to flocculate or agglomerate to form loose clusters of larger particle size.

Particle concentration. The concentration of particles in a dispersoid is often expressed in terms of grains per cubic foot of gas, where 7000 grains = 1 lb In air pollution work, concentrations are sometimes expressed as grams per cubic meter (1 g/m³ = 0 436 grains/ft³). Process dust concentrations normally range from 0 01 to 100 grain/ft³, with 1–10 grain/ft³ being common in process ventilation work. In air-conditioning applications, concentrations of particulate matter will generally range from 0 1 to 10 grains/1000 ft³. In operations such as pneumatic conveying, the concentrations will range from 0.1 to 50 lb solid/lb gas (about 50–20,000 grain/ft³). In fluidized solid systems, concentrations will frequently exceed 1000 lb solid/lb gas.

Figure 1 shows a listing of particle sizes of common materials and related items, as well as methods of size analysis. Figure 2 shows a summary of concentrations of particulates suspended in air.

Particle mechanics. A particle moving in a fluid encounters a frictional resistance

$$F_r = C_D A_p \rho \frac{u_r^2}{2}$$

where F_r is the resisting force in dynes; C_D is the dimensionless drag coefficient; A_p is the area of the particle projected in a plane normal to the direction of motion in cm²; ρ is the fluid density in g/cm³, and u_r is the velocity of the particle relative to the fluid in cm/sec The drag coefficient C_D is a function of the particle shape, the proximity of bounding surfaces, and the Reynolds number $N_{Re}(= D_p u_r \rho/\mu$, where D_p is the particle diameter or other representative dimension in cm; and μ is the fluid viscosity in poises). Drag coefficient data are available for a wide variety of particle shapes.

If a particle suspended in a fluid is acted upon by a force, it will accelerate to a terminal velocity, at which the resisting force due to fluid friction just balances the applied [illegible]

$$u_t = \sqrt{2g_L M_p(\rho_p - \rho)/\rho\rho_p A_p C_D}$$

where u_t is the settling velocity in cm/sec; g_L is the acceleration due to gravity, 980.7 cm²/sec; M_p is the particle mass in grams; ρ_p is particle density in g/cm³; and other terms are as previously defined.

For spherical particles present in dilute concentration, C_D is approximately constant at 0.44 for $N_{Re} > 500$. For $N_{Re} < 1$, C_D is equal to $24/N_{Re}$, and the resisting force is given by

$$F_r = 3\pi\mu u_r D_p$$

which is known as Stokes' law. Corresponding to this, the terminal gravitational settling velocity is

$$u_t = g_L D_p^2(\rho_p - \rho)/18\mu$$

commonly known as Stokes' law of settling, which is usually applicable for particles smaller than 50 μ diameter.

When the particle size approaches the magnitude of the mean free path of the molecules of the suspending fluid, the frictional resistance to motion becomes less than (and the settling velocity becomes greater than) that indicated by the above equations. To correct for this molecular slip flow effect, the resistance to motion calculated from the

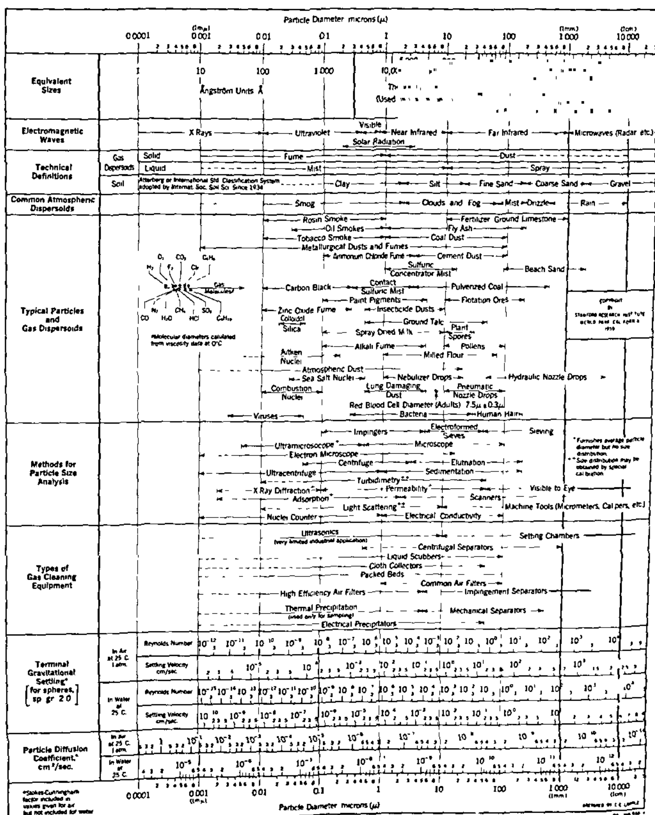


Fig. 1. Characteristics of particles and particle dispersoids (*Stanford Research Institute*)

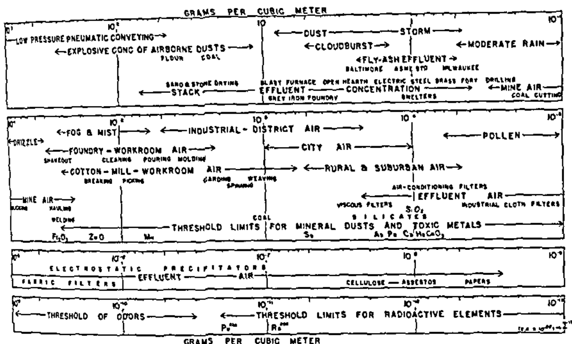


Fig 2 Concentration of particulates in the atmosphere (From M. W. First and P. Drinker, Archives of Ind Hyg and Occupat Medicine, 5 387–388, 1952, by permission American Medical Association)

above equations must be divided by a factor k_m commonly known as the Stokes-Cunningham correction factor, given by

$$k_m = 1 + k_{me}(\lambda/D_p)$$

where k_{me} is a dimensionless constant and λ is the mean free path of the fluid molecules in cm The value k_{me} has been shown experimentally to lie between 1 3 and 2 3 for different gases, particle sizes, and materials. Based on the data of R. A Millikan

$$k_{me} = 1\,644 + 0\,552e^{-(0\,656D_p/\lambda)}$$

where λ is based on *simple* kinetic theory ($\lambda = 3\mu/\rho\bar{v}$), and the mean *molecular* speed $\bar{v}$ is given in cm/sec by

$$\bar{v} = \sqrt{8RT/\pi M}$$

where R is the gas constant, $8\,31 \times 10^7$ ergs/(°C)(g mole), T is the absolute temperature in °K, and M is the molecular weight of the gas in g/g-mole. For particles in gases at atmospheric pressure, the Stokes-Cunningham factor becomes significant for particles smaller than 1 μ At high altitude or low pressure, this factor can become extremely large

Particles suspended in a fluid partake of the molecular motion of the suspending fluid and hence acquire diffusional characteristics analogous to those of the fluid molecules. This random zigzag motion of the particles, commonly known as Brownian motion, is obvious under the microscope for particles smaller than 1 μ The average displacement of a particle in time t is given by

$$\Delta s = \sqrt{4D_v t/\pi}$$

where Δs is the average linear displacement along a given axis, regardless of sign, in cm, D_v is the diffusion coefficient for the particle in cm²/sec; and t is the time in sec The Einstein equation for the diffusion coefficient D_v for spherical particles is

$$D_v = k_m RT/3\pi\mu N_A D_p$$

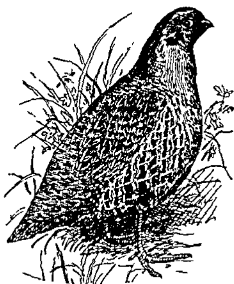
where N_A is Avogadro's number, $6\,023 \times 10^{23}$ molecules/g mole, and other terms are as previously defined This diffusion coefficient may also be used to estimate the migration rate of particles when a concentration gradient exists.

Figure 1 shows settling velocities and diffusion coefficients with particles of various sizes in air and water. *See* AEROSOL, AIR FILTER, COLLOID; DUST AND MIST COLLECTION; FLUIDIZATION OF SOLIDS; SIZE REDUCTION [C.E.L.]

Bibliography: R D Cadle, *Particle Size Determination*, 1955, J M DallaValle, *Micromeritics*, 2d ed, 1948, P Drinker and T. Hatch, *Industrial Dust*, 2d ed, 1954, H L. Green and W. R Lane, *Particulate Clouds Dusts, Smokes, and Mists*, 1957, J J Hermans, *Flow Properties of Disperse Systems*, 1953, Institution of Chemical Engineers and Society of Chemical Industry, *Symposium on Particle Size Analysis*, 1947, C E. Lapple and S. E. Alvis, *Fluid and Particle Mechanics*, 1951; C. H Orr and J M. DallaValle, *Fine Particle Measurement*, 1959, J. H. Perry (ed), *Chemical Engineers' Handbook*, 3d ed, 1950, R W Whytlaw-Gray and H S Patterson, *Smoke*, 1932

Partridge

A name applied to many of the quail-like birds of the family Phasianidae, including the American quails, Old World partridges, and the francolins



The Hungarian partridge, *Perdix perdix*; length to 14 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Partridges are primarily a Northern Hemisphere group, although they are represented elsewhere. The name partridge is generally associated in the United States with two introduced species, the Hungarian partridge, *Perdix perdix*, and the chukar partridge, *Alectoris graeca*. The former is now well established in the southern Canadian plains and in parts of the northern United States. The chukar has been successfully introduced on the talus slopes of the western United States. See GALLIFORMES; QUAIL. [J.D.B.]

Pascal's law

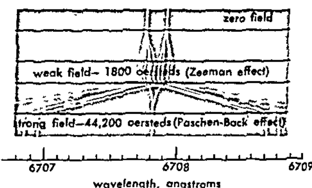
A law of physics which states that a confined fluid transmits externally applied pressure uniformly in all directions. Blaise Pascal, using the mercury-column barometer of Evangelista Torricelli, demonstrated the decrease in atmospheric pressure with increasing height and determined that atmospheric force at a point exerted equal pressure in all directions. More exactly, in a static fluid, force is transmitted at the velocity of sound throughout the fluid. The force acts normal to any surface. The natural phenomenon is the basis of the pneumatic tire, balloon, hydraulic jack, and related devices. See AEROSTAT; HYDROSTATICS. [K.AR.; R.S.R.]

Paschen-Back effect

An effect on spectral lines obtained when the light source is placed in a very strong magnetic field, first explained by F. Paschen and E. Back in 1921. In such a field the anomalous Zeeman effect, which is obtained with weaker fields, changes over to what is, in a first approximation, the normal Zeeman effect (see ZEEMAN EFFECT). The term "very strong field" is a relative one, since the field strength required depends on the particular lines being investigated. It must be strong enough to produce a magnetic splitting that is large compared to the separation of the components of the multiplet. See ATOMIC STRUCTURE AND SPECTRA.

The illustration shows, as an example, the Paschen-Back effect of the red line $2s^2S - 2p^2P$ of lithium. The natural separation of this doublet is very small, only 0.175 A, so that in the field of 44,200 oersteds for which the diagram is drawn the normal Zeeman splitting of 0.929 A greatly exceeds it. The Paschen-Back effect is therefore practically complete, and the resulting pattern is nearly a normal triplet. There is still, however, a residual splitting, which theory gives as two-thirds of the field-free separation. The weak field patterns shown in the figure require a field of only about 1800 oersteds and correspond to the anomalous patterns obtained in the Zeeman effect.

Theory explains the transformation from the Zeeman to the Paschen-Back effect as due to the uncoupling of the orbital and spin vectors L and S by the magnetic field. Whereas in a weak field these vectors are coupled magnetically to form a resultant J, a sufficiently strong field causes them to precess independently about the field direction. The correlation of the energy levels between weak and strong fields follows from the principle that the magnetic field is incapable of changing the angular momentum component along the field; that is, the magnetic quantum number M of a level remains constant for all field strengths. A further rule



Zeeman and Paschen-Back effects of the red lithium doublet.

is that two levels having the same value of M (which in the strong field equals $M_L + M_S$) do not cross each other. These rules lead to the correlation of lines shown in the figure. Certain lines fade out during the transition, as is indicated by the change from a solid to a dotted line, and these are the ones for which the direction of polarization would be altered. [F.A.J.]

Passeriformes

The order of perching birds, containing about 5300 species, or more than half of the living species of birds. It comprises two major divisions, a somewhat more primitive group of 3 suborders, the Eurylaimi, Tyranni, and Menurae, collectively known as suboscines; and the large suborder Passeres or Oscines, the songbirds. Separation of the oscines from the suboscines is based chiefly on the structure of the syrinx and its attached muscles. The Menurae are confined to Australia; the Eurylaimi

that region The largest family of the Tyranni, and the only one extending north of the United States-Mexican border region, is the Tyrannidae, the tyrant flycatchers. A few flycatchers and many cotingas (Cotingidae) and manakins (Pipridae) are highly colorful, but in most Tyranni, including such families as the woodcreepers (Dendrocolaptidae), ovenbirds (Furnariidae), and ant thrushes (Formicariidae), browns, grays, and black predominate in the plumage.

Taxonomy. The division of the Oscines into families and the relationships among groups of families continue to receive much attention from ornithologists. Some families, such as the larks (Alaudidae) and swallows (Hirundinidae), are well marked and present few problems. But among most of the Oscines, family lines are difficult to draw, and relationships of many genera are still uncertain.

Because of uncertainties of their classification, the Oscines have been divided by recent authors into 36-53 families. Some major groups, each of probable monoph...

...en called the primitive insect eaters Within this complex are some families which have evolved entirely or predominantly in the New World, including the wrens (Troglodytidae) and mockingbirds (Mimidae). In both Old and New World are such families as the thrushes (Turdidae), dippers (Cinclidae), and pipits (Motacillidae). The predominantly Old World component of this large group is particularly hard to divide satisfactorily It includes such birds as the Old World flycatchers, warblers, monarchs, babblers, and others.

Although the neotropical passeriform avifauna is dominated by the suboscines, there is also a New World radiation of interrelated oscine groups, also difficult to divide into families Included are such birds as the vireos, wood warblers, honey creepers, tanagers, troupials, cardinal grosbeaks, and emberizine finches

Smaller groups of families include one typified by the nearly world-wide titmouse-nuthatch group, a complex of Australian endemics, a group of predominantly tropical Old World nectar feeders, a group of shrikelike birds, and a large Old World group including such birds as the starlings, weavers, waxbills, and possibly the siskins See AVES [K C P.]

Passive radar

A technique for surveillance, mapping, navigation, and guidance that employs the reception of microwave frequency energy radiated by warm bodies, or reflected from other sources It is similar in principle to infrared systems used for the same purposes See INFRARED RADIATION; MICROWAVE.

Passive radar has some similarity to both passive infrared systems and active (radiating) radar systems Like infrared systems, passive radar emits no energy of its own, and therefore does not give away its position or existence. Because they employ no transmitter, passive radar systems are smaller, lighter, and less complex than active radar. However, an infrared system is smaller, lighter, and less complex than passive radar. The antenna of a passive radar system is comparable to that of an active radar at the same operating frequency; it is larger than that of an infrared system, which operates at higher frequencies Target resolution is inferior to that of active radar and infrared systems However, the ability to discriminate between different targets and backgrounds can be better than either active radar or infrared.

Every object at a temperature above absolute zero radiates electromagnetic energy. Most of it is in the infrared region, but small amounts are emitted throughout the spectrum (*see* HEAT RADIATION). The amount of radiation is determined by the absolute temperature and emissivity of the radiating body, emissivity is defined as the ratio of the object's radiation to that of an equivalent black body at the same absolute temperature. *See* EMISSIVITY.

In addition, objects reflect microwave energy from the sky and other sources. The total radiation received at a receiving antenna is, therefore, the sum of the radiation emitted directly as a result of the object's temperature and the energy reflected into the antenna's beam from the sky and other sources The amount of reflected energy depends on the reflectivity of the body; this provides another way of discriminating between two objects at the same temperature. *See* REFLECTION (ELECTROMAGNETIC RADIATION)

The ability of a passive radar to discriminate between different objects depends primarily on (1) the apparent temperature differential between the objects (this takes into account emissivity and reflectivity), (2) the grazing angle between the antenna beam and the object, (3) antenna polarization, (4) antenna beam width, and (5) the minimum detectable signal level of the receiver. The grazing angle, or angle of incidence, is a factor when viewing smooth objects, such as bodies of water. Polarization is most pronounced at small grazing angles. The effect of polarization varies with the object, providing an additional method of discriminating between objects at the same temperature. *See* RADAR. [P J.K.]

Pasteurization

The application of mild heat, for a specified time, to a liquid food or beverage to enhance its keeping properties and to destroy any harmful microorganisms present For milk, the times and temperatures employed are based upon the thermal tolerance of *Mycobacterium tuberculosis*, one of the most heat resistant of nonsporeforming pathogens A temperature of either 143°F (61.7°C) for 30 min or of 161°F (71 7°C) for 15 sec is employed. Vegetative cells of most bacteria are killed by the heat treatment, but endospores are unaffected *See* MALT BEVERAGE, MILK, WINE.

Pasteurization originated with Louis Pasteur in the 1860s, who demonstrated that spoilage of wine and beer could be prevented by heating these beverages to approximately 135°F (57.2°C) for a few minutes. [C.F.N.]

Patent

Common designation for letters patent, which is a certificate of grant by a government of an exclusive right with respect to an invention for a limited period of time. The exclusive right granted by a patent is the statutory right to make, use, and sell. Portions of those rights deriving naturally from it may be granted separately, as: the rights to use, to make, to have made, to lease, and so forth. Any violation of this exclusive right is an infringement.

An essential substantive condition which must be satisfied before a patent will be granted is the presence of patentable invention or discovery. To be patentable, an invention or discovery must relate to a prescribed category of contribution, such as: process, machine, manufacture, composition of matter, plant, or design. In the United States, there are different classes of patents for different members of these categories.

Mechanical patents. Mechanical patents, which include electrical and electronic patents, are the most familiar; they each have a term beginning upon issue and ending 17 years thereafter. This 17-year class of patent is granted (or issued) for an inventive improvement in a process, manufacture, machine, or composition of matter. A process may, for example, be a method of inducing or promoting a chemical reaction or of producing a desired physical result (differential specific gravity ore separation by flotation). "Manufacture" means any article of manufacture, and includes such diverse items as waveguides, transistors, fishing reels, hammers, buttons, and corks. "Machines" has its broadest conventional meaning, and "composition of matter" includes anything from drugs to alloys. This class of patent affords protection for the principle of the invention or a range of equivalents for doing substantially the same thing in substantially the same way. Thus, mere changes in form, material, inversion, or rearrangement will not avoid infringement of the mechanical patent. A straightforward substitution of one active device for another, as a transistor for an electronic valve, is not such a change as to avoid infringement of a mechanical patent.

Design patent. This class of patent is granted for any new, original, and ornamental design for an article of manufacture. To the extent that shape is determined by functional, rather than ornamental considerations, it is not proper subject matter for a design patent. Unlike the mechanical patent, the design patent may be avoided by a change in proportions or shape, although the essential function may be retained. The design patent is issued for different periods, depending upon the desires of the applicant for patent and the fee paid by him. These terms are 3½, 7, or 14 years.

Plant patents. This class of patent is granted to one who discovers and asexually reproduces any distinct and new variety of plant other than a tuber-propagated plant. The right of exclusion extends only to the asexual propagation of such a plant. The term of the plant patent is 17 years from the date of issue by the Patent Office.

Procedure. The discussion in this and the balance of this article is limited to the mechanical patents, because problems arising with this class of patent have been most thoroughly explored through litigation and because the determinations are applied by extension of reasoning to the other classes of patents in developing their properties and limitations, to the extent they are not expressly disposed of by statute. Letters patent are granted upon the making of written application to the Patent Office. The essential parts of this application are: (1) petition (request that the patent be granted); (2) drawing showing one or more representative embodiments of the invention whenever description of the invention would be aided by its presence; (3*a*) specification (written description of the invention and the manner and process of making and using it in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use it, and setting forth the best mode contemplated by the inventor for carrying out his invention, and the most important element); (3*b*) the claims, which define the scope of the limited term monopoly granted to the patentee, (4) an oath in which the applicant swears that he is the original and first inventor and that the invention was not known or used by others in this country or patented or described in a printed publication in this or a foreign country before his invention or discovery thereof, or patented or described in a printed publication in this or a foreign country or in public use or on sale in this country more than one year prior to the date of his application; and (5) the prescribed filing fee. Owing to the complexities and history of judicial decisions which have developed particular meanings and limitations for these requirements, it is customary to retain an attorney who has been approved or registered by the Patent Office following examination of his qualifications for the purpose of preparing the application and pursuing it through the Patent Office.

Upon receipt of the application, it is examined at the Patent Office by an examiner, who, if conflict exists between the language of the claims and the earlier work of others, points this out in a letter of rejection, specifying the particulars of the basis for rejection. If no conflict is found, the examiner sends the applicant a notice of allowance. In examining the application for the existence of conflicts which would exist if the claims were broad enough to include such earlier work, the examiner consults previous patents and publications. Within six months following the date when the Patent Office mails the rejection, the applicant or his attor-

ner must respond, either by showing that the relationship between the claims and the earlier work of others (prior art) cited by the examiner involves no conflict, or amending the claims to avoid the conflict. This process is repeated until the examiner is satisfied that the applicant is not claiming something which is the invention of another, at which time the notice of allowance is sent to the applicant. The patent is issued upon the payment of the issue fee. The entire process customarily requires from two to five years.

If the examiner and applicant do not ultimately agree upon a proper scope for the claims, the examiner may make a final rejection, which is then subject to review, if desired by the applicant, by a Board of Appeals and, as a further resort, the Court of Customs and Patent Appeals

Invention. Patents are granted for new inventions or discoveries. Discoveries do not include such things as the laws of nature, but do embrace processes, machines, manufactured items, and materials applying such newly discovered laws or properties The necessary attribute of patentable invention is novelty, but it is not a sufficient attribute Invention has never been positively defined, but a number of negative tests are now well recognized To all of these general negative rules, however, there are exceptions whenever something unobvious, not to be anticipated by mere extrapolation, results. Some of the things which generally do not constitute invention are: change in proportions, change in materials [illegible] of known [illegible] cation o [illegible] though there may be invention in the process of purification.

It is generally considered that there are two discernible steps in invention. The first is thinking of it, the second is constructing it. The first is termed conception The second is called reduction to practice.

Conception. The formation in the mind of the inventor of a definite and permanent idea of the complete and operative invention as it is thereafter to be applied in practice is conception It is not merely a perception of what is done, or considered desirable to do, but going beyond this, how it is to be done, in terms of a currently realizable instrumentality or group of instrumentalities. Because conception is a mental act, there must be some external, verifying manifestation in the form of impartation to another, if the act is to be established in later controversy over priority of invention Although oral transmission is adequate in theory, the perishability of the human memory favors the unchanging written word, dated, signed by the inventor, and witnessed by the corroborating party It is this which makes the keeping of written records by the inventor in the course of his work so important, for a failure to have this verified external manifestation of his conception may cost him his patent if the date becomes important and is challenged Because the record is not a proof, but only a document capable of proof, it is important that the recording be in some permanent form in which undetectable alteration is practically impossible, otherwise the weight of the proof will be diminished. Predating of the document is damaging, although a record, made a few days later, of a previously observed event, is valuable, if appearing on its face as such. The parties involved in a record of conception will be tested for verity by cross examination, should the dates ever be challenged. Frequently, the moment of conception is clear only in retrospect, when there is a long sequence of experimental effort directed to attaining the desired result. There is no dramatic thunderclap ushering in the birth of most inventions, hence the desirability of keeping current dated and witnessed records of all work. Such records tend further to corroborate the conception by providing a clue to the entry thought train, and revealing the completeness of understanding by the skill with which its principles are later applied

Conception alone does not give the inventor any vested right in the invention.

Reduction to practice. Two forms of reduction to practice are recognized· actual, or constructive. The filing of a patent application which does not become abandoned is a constructive reduction to practice

Actual reduction to practice requires that the invention be carried out in a physical embodiment demonstrating its practicability In a process or method, this is sufficiently done when the steps are actually carried out to produce the desired result. In a machine or article of manufacture, it is required that there be a construction showing every essential feature of the claimed invention. Practicality is demonstrated by operating the apparatus under the conditions which it is anticipated will be encountered in actual service. For example, the testing of an automobile hot water heater using water from the hot water tap has been held not an actual reduction to practice because some conditions might occur in a motor vehicle which would not be observed in a stationary installation with a heat source of relatively unlimited capacity. Materials are reduced to practice when they are produced, unless utility is not self-evident as, for example, in the case of drug compounds, where it is required to establish that the drug is useful, for the purpose stated, not merely harmless To be effective, the reduction to practice must be by the inventor, or by one acting as his agent, or one who has acquired rights from him

A reduction to practice which results from diligent efforts following conception of the invention gives a vested right in the invention, unless followed by an abandonment. If there is a gap in such diligence, then the effective date of the right to assert ownership of the invention is only that date which can be connected by a continuous train of diligence with the reduction to practice Like conception, reduction to practice must be corroborated by a third party witness, and it is advisable

to have a contemporary written record of what was done and what was observed, accompanied by the date of the observations, to refresh the recollection of the witnesses. The witness must have sufficiently acquainted himself with the internal details of any apparatus to be able later to establish the identity between what was demonstrated and what is sought later to be patented. The diligence required is that which is reasonable under all the circumstances, but must be directed to the reduction to practice of the invention, and to collateral factors. The safest course is to make every effort to reduce to practice consistent with the inventor's physical, intellectual, and financial capacity. For example, if reduction to practice would be clearly within the inventor's financial means, alternative attempts to secure financing could not constitute diligence.

Interferences. When two or more persons are claiming substantially the same invention, a contest to determine priority between the two is instituted by the Patent Office. This proceeding is termed an interference. That party who can carry his reduction to practice back to the earliest date toward and including conception by a continuous train of diligence will prevail in the interference and be awarded the patent, providing those acts have been in the United States. If any or all of the acts were performed outside the United States, the inventor is limited to the date of his first efforts in the United States Exception is made for foreign inventors when they have first filed a corresponding application for patent in a foreign country which is a signatory to the International Convention. In such instance, he will be credited with a date corresponding to his first foreign filing date if the filing in the United States occurs within a year of that date and if he promptly and duly requests this

An applicant may provoke an interference with an issued patent which claims an invention that he believes should rightly belong to him. This is done within 12 months of issue by filing in the Patent Office, as part of a supporting pending or newly filed application for patent, all of the claims which it is sought to contest, and requesting the declaration of an interference.

Following declaration of an interference, the parties are called upon to file preliminary statements under oath setting forth pertinent data surrounding the genesis of the invention and its disclosure to others. When these preliminary statements have been received and approved by the Patent Office, the applicant is notified of the setting of a period of time during which motions may be filed. During this time, access to the adversary's application is permitted, marking one of the few times that the secrecy with which the Patent Office surrounds each patent application is penetrated. During the motion period, various requests for termination of the interference, known as motions, may be presented and set for hearing When an issued patent is involved, there may be no motion for dissolution of the interference upon the ground that the claim or claims in issue is or are unpatentable. After the motions are disposed of, times for taking the testimony are set, during which the parties may take sworn statements from their witnesses before duly qualified officers, which are then filed with the Patent Office, accompanied by any proper exhibits, and used as the basis for the presentation of written arguments, followed by an oral hearing, if the applicant requests. The burden of proof in the interference is on the party who was last to file his application in the Patent Office, and he accordingly has the right to open and close the written and oral presentations. If the junior party does not take testimony during the time allotted to him, then the interference is terminated in favor of the senior party without any testimony being taken by him.

The conduct of an interference proceeding is an arduous and complex matter, being fraught with more technicalities than any other phase of patent law.

Inventor. Only a natural person may be an inventor, as distinguished from a corporation. Inventors may be either sole or joint.

Assignment. Patents and applications for patents have the general attributes of personal property, and interests in them are assignable by instrument in writing. Such assignments will be recorded by the Patent Office upon filing of a request accompanied by copy of the assignment and payment of the proper fee. To be good against subsequent purchasers without notice, the assignment must be recorded within three months from its date, or before the date when such subsequent transfer of rights was made

Witnesses. Corroboration witnesses, as required in connection with the establishment of conception and reduction to practice, must be someone other than the inventor. No joint inventor can serve as a corroborating witness for another joint inventor in connection with the invention which they have jointly made. Beyond this, the rules normally governing witnesses apply, according to which the witness must be qualified intellectually and by training to recognize and to understand the areas in which he testifies

Enforcement. Enforcement of patents is through the Federal Judicial System, action being initiated in the Federal Judicial District where the defendant resides, or where the defendant has committed the alleged act of infringement and has a regular and established place of business. Damages may be awarded, and an injunction granted, prohibiting further infringement by the defendant. If damages are awarded, there can be recovery for a period not longer than six years preceding the filing of the complaint.

When the infringer is the United States Government, or a supplier of the U S. Government operating with the authorization and consent of the Government, the suit must be filed against the Government in the United States Court of Claims.

A patent may become unenforceable through improper use, for example, use as a part of an act

a violation of the antitrust law, or to force the purchase of unpatented parts, materials, or supplies as a condition of securing a license under the patent.

Although no exclusionary rights arise from the filing of an application for patent, the using and selling of goods produced before the issue of the patent may be enjoined, or damages therefore recovered, after the issue of the patent. Under the six year statute of limitations, an action may be maintained on a patent up to six years after its expiration, the accounting being limited in such instance to damages for infringing activities within the six year period.

Licenses Licenses to operate under a patent may be granted, either nonexclusive or exclusive, and may be in writing or may arise as a necessary implication of other actions of the patentee. Except for an exclusive license, licenses are not ordinarily recorded by the Patent Office

Foreign filing. A United States patent is void, if the United States inventor should file an application for the same invention in any country foreign to the United States before six months from the date when he filed in the United States, unless license to do so be first obtained from the Commissioner of Patents Because the United States is a signatory party to the International Convention, applications filed in foreign countries within 12 months of the date when the parent case is filed in the United States are accorded an effective filing date which is the same as the date of filing in the United States The procedural details of foreign filing vary from country to country and from time to time. Generally speaking, the privilege of securing protection for subcombinations (separately useful portions of a larger system) is much more restricted in foreign countries than in the United States

Nuclear and atomic energy. Patents may not be issued for inventions or discoveries useful solely in the utilization of special nuclear material or atomic energy in an atomic weapon, nor does any patent confer rights upon the patentee with respect to such uses As a substitute for the patent incentive for disclosure in this field, there is a mandatory provision in the law that requires anyone making an invention or discovery useful in the production of special nuclear material or atomic energy, in the utilization of such special nuclear materials in the production of an atomic weapon, or in the utilization of atomic energy in an atomic weapon, to report such invention promptly to the Atomic Energy Commission, unless it is earlier described in an application for patent. Awards may then be requested from the Patent Compensation Board of the Atomic Energy Commission as a substitute for the incentive of the patent system.

Aeronautics and astronautics. No patent may be issued for an invention having significant utility in the field of aeronautics or space unless there be filed with the Commissioner of Patents a sworn statement of the facts surrounding the making of the invention and establishing that it was done without any relation to any contract with the National Aeronautics and Space Administration. This is subject to waiver by the Administrator, but in the event of waiver he is required by law to retain a license for the United States and foreign governments.

Marking. A patented product may carry a notice of this fact, including the patent number. The affixation of such notice is of advantage in establishing notice of infringement and fixing the period for which damages may be collected. By statute, false marking is a criminal offense. The marking "Patent Pending" or "Patent applied for" gives no substantive rights, but may give rise to sanction under the above mentioned statute if without foundation in fact.

Foreign patents. The principles guiding most foreign patent systems are essentially the same as those underlying the United States system: the granting of a carefully defined monopoly for a limited term of years in return for a laying open of the invention through letters patent. There are some differences in the classes and terms of patent, and in the nature of subject matter which may be patented Most significant departures are the measuring of the term, in most instances, from the filing date, the imposition of annual taxes increasing each year of the life of the patent, and compulsory licensing [C.V E.]

Pathogen

Any agent capable of causing disease The term pathogen is usually restricted to living agents which include viruses, rickettsia, bacteria, fungi, yeasts, protozoa, helminths, and certain insect larval stages.

Pathogenicity is the ability of an organism to enter a host and cause disease The degree of pathogenicity, that is, the comparative ability to cause disease, is known as virulence The terms pathogenic and nonpathogenic refer to the relative virulence of the organism or its ability to cause disease under certain conditions This ability depends not only upon the properties of the organism but also upon the ability of the host to defend itself (its immunity) and prevent injury The concept of pathogenicity and virulence has no meaning without reference to a specific host For example, gonococcus is capable of causing gonorrhea in man but not in lower animals *See* BACTERIA, FUNGI; GERM; IMMUNITY, MYCOLOGY, MEDICAL, MYIASIS, PARASITOLOGY, MEDICAL, PLANT DISEASE; PLANT VIRUS, PROTOZOA, RICKETTSIOSES; VIRULENCE; VIRUS. [D N.L.]

Pathogen (soil)

Any microbe in the soil that causes disease in man, animals, and plants While a majority of these organisms live in the soil for only a short period of time, a few, such as *Bacillus anthracis*, can survive for considerable periods.

Human and animal pathogens. Soil is unsuited for the survival of the great majority of microor-

ganisms causing disease in humans and animals. With few exceptions, microbes pathogenic for man or animals are not indigenous to soil. Many disease organisms find their way into soil in the excreta of their hosts or their remains or reach the soil through contaminated dust or surface waters. In most cases they find soil an uncongenial habitat and are able to remain viable for only short periods. The soil is not a serious source of epidemics of infectious diseases.

Most disease-producing bacteria find soil deficient in the special nutrients to which they have become adapted as parasites and are thus unable to compete successfully with the normal soil microorganisms in utilizing the available food. Furthermore, the antagonistic action exerted by many species of soil bacteria, actinomycetes, and fungi is a factor in the inhibition and eventual elimination of most pathogens. Survival is shorter in acid and alkaline soils than in more neutral soils, in loam or clay soils than in sandy soils, and in warmer than in cooler soils. In sterilized soil, devoid of natural antagonists, pathogens remain viable longer than in natural soils.

Bacteria causing typhoid and paratyphoid fevers, dysentery and cholera, as well as many other pathogens, such as those responsible for pneumonia, plague, and streptococcal and staphylococcal infections, survive for only brief periods, measured in days. Typhoid bacteria have been found to disappear within a week after being added to soil. Tubercle bacteria persist somewhat longer. The bovine strain of *Mycobacterium tuberculosis*, present in cow feces on pasture land, has been found to survive for periods varying from 2 months in summer to 5 months in winter. *See* BACTERIOLOGY, MEDICAL.

Certain sporeforming pathogens remain viable for long periods in soil or persist indefinitely. The anthrax organism, in areas where the disease is prevalent, contaminates ground used by infected animals and may remain alive for 12 years or more to reinfect animals, particularly cattle and sheep feeding on such ground. Anaerobic sporeformers, such as those causing tetanus and gas gangrene, are widely distributed in soils. The concentration of spores is not high and individual samples may not reveal the organisms. The same holds for *Clostridium botulinum* which, though not responsible for an infection, is able to produce a powerful toxin when allowed to grow in foods. Anaerobic pathogenic sporeformers occur more commonly in virgin than in cultivated soils.

Plant pathogens. The soil also harbors organisms that cause plant diseases. Bacteria are responsible for such diseases as soft rots of vegetables and certain leaf spots and galls. Other diseases, such as potato scab, are caused by actinomycetes. Fungi comprise the most destructive plant pathogens and cause a variety of diseases, including root rots of cereals and other crops, various wilts, blights, and mildews. The pathogenic fungi may be divided broadly into either root-inhabiting fungi, the specialized parasites which disappear rapidly in the absence of the host plant, or facultative parasites which are able to survive as saprophytes in soil. Certain nematodes may cause damage to the roots of many crops, not only by injuring the root tissue, but by facilitating infection by other parasites. The ability of soil-borne microorganisms to cause infection is influenced by such factors as soil texture, reaction, moisture, temperature, fertilizer treatment, and crop sequence. *See* PLANT DISEASE; SOIL MICROBIOLOGY. [A.C.L.]

Pathology

The study of the causes, alterations, and mechanisms related to disease. In the broad sense, pathology includes the study of both plant and animal diseases. Human pathology is the more restricted, medically applicable field which derives its subject matter from whatever area is likely to produce useful information.

Cellular and histologic pathology approach disease at the microscopic level whereas gross pathology deals with the visible alterations seen in tissues and organs. Clinical pathology gathers methods and principles from chemistry, microbiology, and other disciplines for use at the bedside or in the hospital laboratory as diagnostic aids.

As medicine has expanded, corresponding fields of pathology have developed into subspecialties, such as surgical pathology and neuropathology. Experimental pathology attempts to study disease mechanisms under controlled conditions often utilizing highly technical methods. General pathology covers all of these areas, although in less detail, and serves in medical education as the means of introducing the subject of disease to medical students

Disease. The term disease has many meanings To most people it represents a fairly specific sequence of changes characteristic of measles, arthritis, and other illnesses. To the pathologist, disease is a concept, not a thing. It represents the reaction of the body to stimuli which tend to alter either structure or function. The results of such alteration is really the illness, be it specific or vague in nature The cause of the reaction is often less clear but involves the stimulus, or etiologic agent, and then a complex series of reactions between the original stimulus and the besieged tissues. Several outcomes are possible. The stimulus may be overcome without causing any change in structure or function, that is, no illness ensues. The stimulus may cause a reaction which temporarily produces altered form of function which is then followed by a successful return to normal. A reaction may develop in which there is an unsuccessful or only partially successful outcome so that abnormalities in form or function occur. Depending on the location and severity of these changes they may produce results which range from insignificant to those incompatible with life.

Tissue reaction. As is well known, the same stimulus may precipitate a different series of re-

ations in various tissues or organs. No two persons react in identical ways to the same stimuli; this principle, while permitting broad patterns to be drawn for the typical case, explains some of the inherent difficulties in diagnosis and elucidation of mechanisms of alterations Finally, many other factors may influence the situation, such as those of heredity, environment, age, and sex.

Despite the apparent complexity of alteration in form or function, it is believed that these changes are brought about by basically simple mechanisms which must be found The basis for this belief lies in the fact that a certain cell or tissue, because of its inherent nature, can react to injury in only a limited number of ways In general, the more specialized the tissue, as for example that of the nervous system, the less the range of reactivity. However, each major category of tissue (nervous, muscle, epithelial, and connective) has certain characteristics and capabilities of reaction Some combinations of these cells react actively by proliferation, others react by changing or adapting in some way, and still others degenerate or die

Because of the interdependence of various tissues, reaction in one part of the body may often cause secondary changes elsewhere The integrative and coordinating functions of the nervous system and the fluid communication system of the blood and lymphatic systems [illegible]

[illegible]

[illegible] and L. Pasteur, the principal emphasis was placed on the study of etiologic agents, especially the microorganisms *Bacteria*, spirochetes, fungi, parasites, rickettsia, and viruses continue to be objects of intensive *investigation* regarding their relationship to illness. *In addition*, nonliving agents may cause alterations of tissue. Common examples include radiant energy, such as heat, x-ray, ultraviolet, and *other rays*, and also physical forces or chemical substances. For the most part, these are extrinsic agents, that is, external Far more subtle are the intrinsic *stimuli*, or lack of stimuli, which cause defects in development and metabolism

Classification of changes. *Certain* broad categories of alteration may be *mentioned. Congenital* defects and errors in *development arise* from a host of extrinsic and intrinsic *factors such* as heredity, maternal diseases, embryonic inadequacies, and intrauterine diseases

Inflammations are reactions to injury characterized by edema, congestion, exudation, and various types of proliferation of cells These *may be* seen in response to stimulation by many etiologic *agents*, such as microorganisms, those of sensitivity reactions, and physical or chemical injury. Such inflammations may be quite specific, as in *tuberculosis*, or may be *nonspecific*, that is, without features that indicate the agent, as in the common cold and related conditions

The disorders marked by neoplasia, the development of new growths, are of great *medical* importance. Malignancies are neoplasias in which growth is uncontrolled. If treatment is not effective, malignancies eventually will cause death The other, more common category of neoplasias is that of benign growths. These consist of increased numbers of essentially normal cells which do not spread, or metastasize, and which do not directly cause death, although their physical presence in a vital structure may produce fatal results There is a small group of neoplasias which are borderline between malignancies and benign tumors.

Many pathologic processes are *marked* by degenerative changes in organs or tissues, the causes of which may be unknown or only suspected, as in the case of arteriosclerosis. Sometimes these degenerative changes are related to hereditary factors or to specific agents to which a cell has no effective response

Faulty metabolism and nutritional defects often produce changes in form or function which are quite specific, as for example in gout and the vitamin deficiencies. In many cases, the actual cause of such derangements is obscure and these can only be grouped on the basis of the broad type of mechanism displayed.

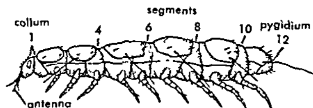
Many disorders defy classification in even a heterogeneous list as given above, either because they show characteristics of different alterations, or because the etiology is unknown, the alterations are subtle, and the mechanisms are obscure

Areas of investigation. In summary, pathology considers the factors which are known or suspected to be related to alterations of form or function. It attempts to produce or implicate an etiologic agent and finally seeks to delineate the mechanisms involved in producing sequential changes in the body.

Present-day pathologic investigation covers all fields of medicine and related *sciences Intensive* efforts are being made to work from the illness to the submicroscopic changes in molecular relationships that must ultimately produce the disorders which are seen microscopically, grossly, and clinically Perhaps eventually the abstract concept of disease can be related to electrophysical *changes* in basic cellular components, because all form and function finally depends upon atomic aggregations and energy exchanges *See* BACTERIOLOGY, MEDICAL, CLINICAL PATHOLOGY; DISEASE, GERONTOLOGY, MYCOLOGY, MEDICAL; ONCOLOGY; PARASITOLOGY; PLANT DISEASE. [E.G.ST]

Pauropoda

A class, and perhaps the most obscure group, of the Myriapoda They are pale creatures, no longer than a millimeter or two, inhabiting damp situations in leaf litter, under bark, stones and *debris, in* humus and similar detritus. Their size and retiring habits suggest rarity, this, however, has been refuted by recent studies that show them to occur in huge *numbers in suitable* habitats. Apparently very *widely distributed* as a class, they have been undiscovered only in deserts and in the arctic *and* antarctic regions Like millipedes, they are progo-



Pauropoda, Pauropus silvaticus Tiegs, fully extended adult. Greatly enlarged *(After O. W. Tiegs, from R. E. Snodgrass, A Textbook of Arthropod Anatomy, Cornell Univ. Press, 1952)*

neate, have one pair of maxillae, and their trunk segments display a certain degree of amalgamation manifest in the tendency to form diplotergites, but not diplosegments. Their peculiar bifurcate antennae and adult complement of 12 trunk segments with 9 pairs of functional legs are distinctive within the myriapod complex. All pauropods lack eyes, spiracles, tracheae, and a circulatory system.

What they eat and how they live is virtually unknown. They probably subsist largely or entirely upon decaying plant and animal matter; some may consume microscopic animals. Studies based upon a few forms reveal an anamorphic ontogeny similar to that of diplopods. The young pauropod hatches with three pairs of legs, then molts successively until the adult complement of nine is attained.

The class currently consists of 2 families with less than 10 genera; there are probably fewer than 60 species known. *See* MANDIBULATA; MYRIAPODA. [R.E.C.]

Pavement

An artificial surface laid over the ground to facilitate travel. In this article only road pavements will be discussed. The engineering involved is closely similar to that for airport surfacing, another major type of pavement. A pavement's ability to support loads depends primarily upon the magnitude of the load, how often it is applied, the supporting power of the soil underneath, and the type and thickness of the pavement structure. Before he can calculate the necessary thickness of a pavement, the engineer must know the volume, type, and weight of the traffic (the traffic load) and the physical characteristics of the underlying soil.

Traffic load. Traffic data can readily be obtained. Traffic is counted to learn the total volume and the proportion of heavy vehicles. Loadometer scales are placed next to the road and trucks are weighed, front and rear wheels separately. Traffic trends are studied. Such data provide a basis for estimating the total volume of traffic to be carried on a pavement during its service life. They also permit an estimate of the magnitude and frequency of the expected load. *See* TRAFFIC ENGINEERING.

Base and surface courses of pavements are designed to withstand many applications of load over a prolonged period, in some cases up to 30 years. In structural design it is also necessary to give consideration to the direction of traffic. For example, hauling from a mine to a nearby railroad siding may result in a high percentage of heavily loaded trucks on the in-bound lanes and a low percentage on the return lanes.

In general, the larger the volume of heavy vehicles on a highway, the greater the structural capacity required in the pavement. But equally as important as the volume of heavy vehicles are the magnitude of the applied loads and the conditions that will influence the effect of those loads on the pavement. Under the action of vehicular traffic, the surface of a pavement is subjected to a series of highly concentrated forces applied through the wheels of the vehicle. These forces exert an influence throughout the depth of the pavement.

The applied loads vary considerably, depending upon the number and spacing of the wheels of each vehicle, the gross weight of the vehicle, and the distribution of that weight among the axles. Two vehicles of the same gross weight may thus differ widely in the wheel loads applied to the highway. A relatively small truck may cause a load of higher unit stress than a larger vehicle with larger tires and more axles. Consideration of the actual wheel loads has become increasingly important with the large increase in the volume of heavy vehicles on most highways.

All highways, regardless of their design, have some limit to their ability to support the frequent application of a heavy load. A large number of load applications can be supported by a given pavement if the load does not exceed a particular magnitude, but once this magnitude is exceeded, distress and failure of the pavement becomes increas evident. The ability of the pavement to suppo load is also influenced by weather. Definite limitations are imposed by law on many high Further limitations are often imposed on sp highways of lighter design, especially during conditions as spring freezing and thawing.

In the preceding discussion wheel loads been treated as static loads. For adequate d analysis the effects of dynamic loads must be uated. The vertical force exerted on the pave by a moving wheel may be considered to b sum of the static weight of the wheel and th pact or dynamic force from the wheel's move over irregularities in the pavement surface. static load will be a constant factor except a movement of the vehicle along the highway s or oscillates the load.

An exceedingly variable factor, the dyn force generated by a moving wheel depends (1) the magnitude of the static wheel load, (2) operating speed of the vehicle, (3) the type, and cushioning properties of the tire equipn and (4) the smoothness of the pavement. Ar crease in static wheel load or pavement rough a decrease in the cushioning qualities of the t and, within limits, an increase in vehicle speed result in increased dynamic forces.

The importance of the foregoing variables factors has long been recognized by design eng neers. The difficulty has been to express the data i

terms that could be rationally applied to design formula and then correlated with the performance of foundation soils and materials used to construct the pavement. The problem has not been given a rigorous mathematical solution but rather has depended largely upon field observations under actual operating conditions

Several methods of load evaluation have been developed and used by various highway agencies [illegible] methods are more or [illegible]

dicate changing conditions of traffic, climate, or soil performance No universal method of load evaluation has been developed so far and indeed cannot be until a method is devised that can be readily adjusted when other factors influencing the performance of a pavement structure change.

Methods of evaluation now in use include (1) numerical count method, in which the number of vehicles using a particular highway is actually counted and the weight of various vehicles listed as light, medium, heavy, and extra heavy; (2) wheel load method, in which factors based upon the actual weight of the wheel load are used; (3) load frequency method, in which the wheel load weights are combined with the volumetric count of the commercial vehicles; and (4) equivalent wheel load method, in which destructive effects of the actual wheel loads being applied to a pavement are expressed in terms of a standard wheel load.

Evaluation of subgrade. Factors that must be considered in evaluating the ability of the underlying soil or subgrade to support the pavement include type of soil, such as loam or clay, gradation and variation in particle size; strength or bearing value; modulus of deformation; and swell or volume change characteristics and related properties

Measurement of the supporting power of the subgrade presents numerous difficulties. Tests sometimes used include the plate bearing test, the direct shear method, the triaxial compression test, and the bearing ratio procedure. See SOIL MECHANICS

Some soil types are unsuitable for supporting pavements because they have low bearing values or undergo changes in volume with variation in moisture content. For example, it is desirable to excavate peat and muck along the road and replace it with soil of higher bearing value

Rigid and flexible pavements Once the grading operation has been completed and the subgrade compacted, construction of the pavement can begin. Pavements are either flexible or rigid Flexible pavements have less resistance to bending than do rigid pavements Both types can be designed to withstand heavy traffic. Selection of the type of pavement will depend, among other things, upon (1) estimated construction costs, (2) experience of the highway agency doing the work with each of the two types, (3) availability of contractors experienced in building each type, (4) anticipated yearly maintenance costs, and (5) experience of the owner in maintenance of each type.

Flexible pavements. Flexible paving mixtures are composed of aggregate (sand, gravel, or crushed stone) and bituminous material The latter consists of asphalt products, which are obtained from natural asphalt products or are produced from petroleum; and tar products, which are secured in the manufacture of gas or coke from bituminous coal or in the manufacture of carburetted water-gas from petroleum distillates. Structural strength of a bituminous pavement is almost wholly dependent upon the aggregate, which constitutes a high percentage of the volume of the mixture and forms the structure that carries the wheel load stresses to the base layers. The bituminous material cements the aggregate particles into a compact mass with enough plasticity to absorb shock and jar; it also fills the voids in the aggregate, waterproofing the pavement.

Among the many types of bituminous surface used are surface treatments, penetration macadams, road mixes, and plant mixes, as well as variations of these.

With surface treatment a liquid bituminous material is applied over a previously prepared aggregate base

In building penetration-macadam pavement, a base, usually of crushed stone or gravel, is constructed in layers and firmly compacted by rolling Often during the rolling, water is applied to make what is termed a water-bound base, or macadam. Then, the keyed and wedged fragments in the upper portion of the base are bonded in place by working alternate applications of bituminous material and choke stone into the surface voids.

With road mix designs a base is constructed, and a layer of aggregate and bituminous material, mixed on the road with a motor grader, is then uniformly spread and compacted with rollers.

When a plant-mix design is used, construction of a base is also necessary, but the aggregate and bituminous materials are mixed at a central plant, trucked to the job, and then spread or placed with a paving machine and compacted With the plant-mix procedure more accurate mixing is possible.

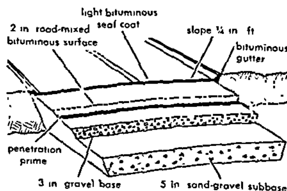
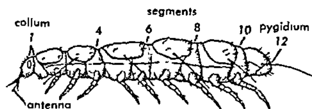


Fig 1 Flexible pavement design for a city collector street with maximum traffic load of 5 tons per axle Right-of-way is 60 ft wide Berms or boulevards at sides are sloped to drain toward the street.



Pauropoda, *Pauropus silvaticus* Tiegs, fully extended adult. Greatly enlarged. (After O. W. Tiegs, from R. E. Snodgrass, *A Textbook of Arthropod Anatomy*, Cornell Univ. Press, 1952)

neate, have one pair of maxillae, and their trunk segments display a certain degree of amalgamation manifest in the tendency to form diplotergites, but not diplosegments Their peculiar bifurcate antennae and adult complement of 12 trunk segments with 9 pairs of functional legs are distinctive within the myriapod complex. All pauropods lack eyes, spiracles, tracheae, and a circulatory system.

What they eat and how they live is virtually unknown. They probably subsist largely or entirely upon decaying plant and animal matter; some may consume microscopic animals. Studies based upon a few forms reveal an anamorphic ontogeny similar to that of diplopods. The young pauropod hatches with three pairs of legs, then molts successively until the adult complement of nine is attained

The class currently consists of 2 families with less than 10 genera; there are probably fewer than 60 species known *See* MANDIBULATA; MYRIAPODA [R E.C.]

Pavement

An artificial surface laid over the ground to facilitate travel. In this article only road pavements will be discussed The engineering involved is closely similar to that for airport surfacing, another major type of pavement A pavement's ability to support loads depends primarily upon the magnitude of the load, how often it is applied, the supporting power of the soil underneath, and the type and thickness of the pavement structure. Before he can calculate the necessary thickness of a pavement, the engineer must know the volume, type, and weight of the traffic (the traffic load) and the physical characteristics of the underlying soil.

Traffic load. Traffic data can readily be obtained. Traffic is counted to learn the total volume and the proportion of heavy vehicles Loadometer scales are placed next to the road and trucks are weighed, front and rear wheels separately. Traffic trends are studied Such data provide a basis for estimating the total volume of traffic to be carried on a pavement during its service life. They also permit an estimate of the magnitude and frequency of the expected load. *See* TRAFFIC ENGINEERING.

Base and surface courses of pavements are designed to withstand many applications of load over a prolonged period, in some cases up to 30 years. In structural design it is also necessary to give consideration to the direction of traffic. For example, hauling from a mine to a nearby railroad siding may result in a high percentage of heavily loaded trucks on the in-bound lanes and a low percentage on the return lanes.

In general, the larger the volume of heavy vehicles on a highway, the greater the structural capacity required in the pavement. But equally as important as the volume of heavy vehicles are the magnitude of the applied loads and the conditions that will influence the effect of those loads on the pavement. Under the action of vehicular traffic, the surface of a pavement is subjected to a series of highly concentrated forces applied through the wheels of the vehicle. These forces exert an influence throughout the depth of the pavement.

The applied loads vary considerably, depending upon the number and spacing of the wheels of each vehicle, the gross weight of the vehicle, and the distribution of that weight among the axles. Two vehicles of the same gross weight may thus differ widely in the wheel loads applied to the highway. A relatively small truck may cause a load of higher unit stress than a larger vehicle with larger tires and more axles. Consideration of the actual wheel loads has become increasingly important with the large increase in the volume of heavy vehicles on most highways.

All highways, regardless of their design, have some limit to their ability to support the frequent application of a heavy load. A large number of load applications can be supported by a given pavement if the load does not exceed a particular magnitude, but once this magnitude is exceeded, distress and failure of the pavement becomes increasingly evident. The ability of the pavement to support the load is also influenced by weather. Definite load limitations are imposed by law on many highways. Further limitations are often imposed on specific highways of lighter design, especially during such conditions as spring freezing and thawing

In the preceding discussion wheel loads have been treated as static loads For adequate design analysis the effects of dynamic loads must be evaluated. The vertical force exerted on the pavement by a moving wheel may be considered to be the sum of the static weight of the wheel and the impact or dynamic force from the wheel's movement over irregularities in the pavement surface. The static load will be a constant factor except as the movement of the vehicle along the highway sways or oscillates the load.

An exceedingly variable factor, the dynamic force generated by a moving wheel depends upon (1) the magnitude of the static wheel load, (2) the operating speed of the vehicle, (3) the type, size, and cushioning properties of the tire equipment, and (4) the smoothness of the pavement An increase in static wheel load or pavement roughness, a decrease in the cushioning qualities of the tires, and, within limits, an increase in vehicle speed all result in increased dynamic forces.

The importance of the foregoing variables and factors has long been recognized by design engineers. The difficulty has been to express the data in

terms that could be rationally applied to design formula and then correlated with the performance of foundation soils and materials used to construct the pavement. The problem has not been given a rigorous mathematical solution but rather has depended largely upon field observations under actual operating conditions.

Several methods of load evaluation have been developed and used by various highway agencies and organizations. All of the methods are more or less empirical in approach and thus are subject to revision and adjustment when field observations indicate changing conditions of traffic, climate, or soil performance. No universal method of load evaluation has been developed so far and indeed cannot be until a method is devised that can be readily adjusted when other factors influencing the performance of a pavement structure change.

Methods of evaluation now in use include (1) numerical count method, in which the number of vehicles using a particular highway is actually counted and the weight of various vehicles listed as light, medium, heavy, and extra heavy; (2) wheel load method, in which factors based upon the actual weight of the wheel load are used, (3) load frequency method, in which the wheel load weights are combined with the volumetric count of the commercial vehicles, and (4) equivalent wheel load method, in which destructive effects of the actual wheel loads being applied to a pavement are expressed in terms of a standard wheel load

Evaluation of subgrade. Factors that must be considered in evaluating the ability of the underlying soil or subgrade to support the pavement include type of soil, such as loam or clay; gradation and variation in particle size; strength or bearing value; modulus of deformation, and swell or volume change characteristics and related properties.

Measurement of the supporting power of the subgrade presents numerous difficulties. Tests sometimes used include the plate bearing test, the direct shear method, the triaxial compression test, and the bearing ratio procedure. See SOIL MECHANICS.

Some soil types are unsuitable for supporting pavements because they have low bearing values or undergo changes in volume with variation in moisture content. For example, it is desirable to excavate peat and muck along the road and replace it with soil of higher bearing value.

Rigid and flexible pavements. Once the grading operation has been completed and the subgrade compacted, construction of the pavement can begin. Pavements are either flexible or rigid. Flexible pavements have less resistance to bending than do rigid pavements. Both types can be designed to withstand heavy traffic. Selection of the type of pavement will depend, among other things, upon (1) estimated construction costs, (2) experience of the highway agency doing the work with each of the two types, (3) availability of contractors experienced in building each type, (4) anticipated yearly maintenance costs, and (5) experience of the owner in maintenance of each type

Flexible pavements. Flexible paving mixtures are composed of aggregate (sand, gravel, or crushed stone) and bituminous material. The latter consists of asphalt products, which are obtained from natural asphalt products or are produced from petroleum; and tar products, which are secured in the manufacture of gas or coke from bituminous coal or in the manufacture of carburetted water-gas from petroleum distillates. Structural strength of a bituminous pavement is almost wholly dependent upon the aggregate, which constitutes a high percentage of the volume of the mixture and forms the structure that carries the wheel load stresses to the base layers The bituminous material cements the aggregate particles into a compact mass with enough plasticity to absorb shock and jar; it also fills the voids in the aggregate, waterproofing the pavement.

Among the many types of bituminous surface used are surface treatments, penetration macadams, road mixes, and plant mixes, as well as variations of these.

With surface treatment a liquid bituminous material is applied over a previously prepared aggregate base.

In building penetration-macadam pavement, a base, usually of crushed stone or gravel, is constructed in layers and firmly compacted by rolling. Often during the rolling, water is applied to make what is termed a water-bound base, or macadam. Then, the keyed and wedged fragments in the upper portion of the base are bonded in place by working alternate applications of bituminous material and choke stone into the surface voids.

With road-mix designs a base is constructed, and a layer of aggregate and bituminous material, mixed on the road with a motor grader, is then uniformly spread and compacted with rollers

When a plant-mix design is used, construction of a base is also necessary, but the aggregate and bituminous materials are mixed at a central plant, trucked to the job, and then spread or placed with a paving machine and compacted With the plant-mix procedure more accurate mixing is possible.

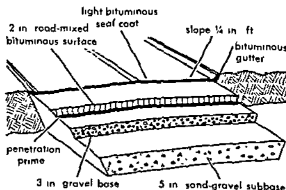


Fig 1. Flexible pavement design for a city collector street with maximum traffic load of 5 tons per axle Right-of-way is 60 ft wide Berms or boulevards at sides are sloped to drain toward the street.

For flexible pavements designed to support heavy loads a subbase built of materials similar to those of the base but of poorer quality may also be used Thickness of the wearing surface, the base, and the subbase depends upon the design load. A typical flexible pavement design for light loads or a 5-ton axle loading is shown in Fig. 1.

Rigid pavements Coarse aggregate, fine aggregate, and portland cement are mixed with clean acid-free water to produce the concrete used for rigid pavements. The coarse aggregate may consist of coarse gravel or crushed stone and the fine aggregate of sand or crushed-stone screenings. The thickness of the pavement slab may vary from 6 in. for light traffic to 18 in. or more for airport pavements accommodating heavy aircraft A layer of granular material such as sand, sandy gravel, or slag is generally used as a subbase under the concrete slab to prevent frost heave and to increase the supporting power of the underlying soil. A rigid pavement is shown in cross section in Fig. 2.

No steel reinforcement is used with bituminous or flexible pavements. With rigid pavements, especially those designed for heavy loads, reinforcement is often used to strengthen the pavement and to prevent cracking. The reinforcement may consist of welded wire fabric or bar mats assembled by tying transverse and longitudinal steel rods together at their point of intersection The reinforcement is usually placed about 2 in. below the upper surface of the concrete slab.

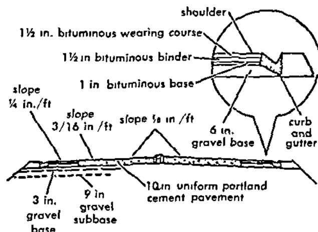


Fig 2. The main roadway of this design is rigid and the shoulders are flexible. The design is for maximum loads of 9 tons per axle

In constructing rigid pavement a longitudinal joint is used between adjacent lanes. Transverse joints, such as expansion and contraction joints to prevent cracking of the pavement when the temperature changes, may also be included. With flexible pavement no joints are used.

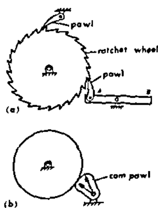
Research. Much remains to be learned regarding the performance and length of service of various types of pavement under different traffic, foundation, and climatic conditions. Probably the most comprehensive experimental pavement study ever undertaken is now in progress near Ottawa, Illinois. Costing several million dollars, this project is being financed jointly by all the state highway departments, several Federal agencies, and numerous private organizations Pavements of different types and thicknesses are being investigated. *See* CONCRETE, HIGHWAY ENGINEERING. [A.N.C.]



Fig 3. A rigid pavement reinforced by welded steel wire mesh Underneath the mesh are 6 in. of concrete. A 3-in layer is being laid on top.

Pawl

The driving link or holding link of a ratchet mechanism. *In the figure, the driving pawl at A, forced* upward by lever *B*, engages the teeth of the ratchet wheel and rotates it counterclockwise. Holding pawl *C* prevents clockwise rotation of the wheel when the pawl at *A* is not engaged. Driving and holding pawls likewise engage rack teeth on the plunger of a ratchet lifting jack, such as those supplied with automobiles. A ratchet wheel with a holding pawl only, acting as a safety brake, is fastened to the drum of a capstan, winch, or other



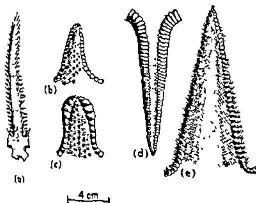
Two types of pawl (a) ratchet mechanism, (b) cam pawl

powered hoisting device A cam pawl, by a wedging action, prevents the wheel from turning clockwise, while permitting free counterclockwise rotation This principle is used in the automobile hill holder, which prevents its rolling backward. In England, a pawl is called also a click.

The spacing mechanism of a typewriter, although frequently termed an escapement, is properly a ratchet device, in which a holding pawl is withdrawn from a spring loaded rack to allow movement of the carriage, while an arresting pawl is introduced momentarily to permit the holding pawl to engage the next tooth of the rack. [E.S.F.]

Paxillosa

A suborder of Phanerozonia with pointed tube-feet which lack suckers and if both series of marginals are present, the upper ones lie directly over the lower ones Paxillae usually cover the upper surface of the body. Of the three included families the



Representative Paxillosa (a) *Luidia neozelanica*. (b) *Asterodon miliaris* (c) *Asterodon dilatatus* (d) *Pseudarchaster abernethyi* (e) *Persephonaster neozelanicus*.

Porcellanasteridae are notable as essentially deep-water forms; the genera *Porcellanaster*, *Eremicaster*, *Thoracaster*, and *Styracaster* have all been found in trenches deeper than 4 miles. The families Astropectinidae and Luidiidae occur in all seas from tidal level downward Luidiidae are distinguished by their straplike arms which lack the upper marginals See PHANEROZONIA. [H.B.F.]

Pea

An edible legume known to be of ancient origin. Introduced into China from Persia about 400 A.D., peas later became one of the early sources of food in Britain, were brought to America with the first colonists, and have been an important part of the American diet since that time. Peas are grown in every part of the United States and are one of the most popular garden vegetables Commercial growing is limited mainly to the Northern States where the growing season is relatively cool.

Because of their commercial importance, two types of garden peas are discussed in this article; both were developed from the species *Pisum*



Typical wrinkled-seed type of pea showing fruit (pod) containing seeds (peas) and leaves and vines. (Washburn Wilson Seed Company, Moscow, Idaho)

sativum These are the wrinkled-seed type which is harvested green and used as a fresh vegetable or for freezing or canning, and the smooth-seed type which may be harvested green for canning and freezing or harvested dry and used for food, either split or whole.

Propagation. Peas are propagated from seed.

that are grown for cover crops have colored flowers Peas should be planted in a well-pulverized, firm seed bed in contact with a good supply of moisture. The date for planting varies with the area and with climatic conditions. For example, in the Palouse area of Washington and Idaho, it ordinarily extends from April 15 to May 10 (most farmers prefer the earlier date). The peas are seeded with an ordinary grain drill in the distribution of 120–160 lb/acre

Harvesting. Green peas are harvested when a processed product of highest quality can be obtained This determination is made by a device called a tenderometer, which, by means of a series of knives, measures the pressure required to cut through a sample of green peas. As the pea matures beyond the optimum point, it becomes less tender and yields a lower-quality canned product.

The harvesting of green peas consists of mowing, windrowing, and loading the vines into trucks which haul them to a viner This machine separates the peas from the vines and pods The last two are made into an excellent high-protein silage feed. The peas are placed in boxes of approximately 25-lb capacity and hauled to the canning or freezing plant where they are cleaned, washed, and, in the case of peas

for canning, screened to six different sizes. They are then hand picked from moving belts, blanched, cooled, and sent to the canner or freezer.

Dry peas are harvested at maturity when the moisture content drops to 14% or lower. Ordinary grain-harvesting combines with minor adjustments are used. The peas are taken to a warehouse where a representative sample is drawn and sent to a federally supervised pea-grading laboratory. The sample is analyzed and a certificate which states the per cent of dockage and defective peas is issued. Peas are sold by the farmer to the warehouseman on the basis of this certificate. The dockage is removed from the peas by scalper machines, and weevily peas are separated by gravity machines. Many warehouses use electronic sorting machines to pick out damaged peas, foreign material, and peas of other classes from seed peas. Dry commercial peas are sold as whole peas or processed further by splitting. The splitting machine separates the two cotyledons and removes the seed coat, leaving an attractive food product. Processed peas, both whole and split, are sampled and graded according to standards set up by the Grain Division of the Agricultural Marketing Service of the U.S. Department of Agriculture A grade certificate is issued, placing the peas in U.S. No 1, 2, 3, or sample grade, thus determining the relative market values.

Production. Wisconsin, Washington, Minnesota, Oregon, Illinois, New York, Pennsylvania, Utah, and Id[illegible]
green [illegible]
the lea[illegible]
Califor[illegible], [illegible], [illegible], [illegible] Dakota,

Pea production in the United States, 1946–1957 averages

State	Harvest, acres	Production, 100-lb bags	Value, dollars
Dry peas*			
Washington	157,000	1,841,000	8,561,000
Idaho	104,000	1,248,000	5,803,000
Oregon	12,000	123,000	571,000
California	11,000	106,000	492,000
Colorado	11,000	93,000	432,000
Montana	7,000	81,000	376,000
North Dakota	5,000	58,000	269,000
Wyoming	4,000	57,000	265,000
Minnesota	4,000	42,000	195,000
United States	317,000	3,651,000	17,077,000
Green peas†			
Wisconsin	127,000	2,629,000	11,181,000
Washington	62,000	1,478,000	6,499,000
Oregon	55,000	1,108,000	4,731,000
Minnesota	52,000	969,000	4,646,000
New York	24,000	436,000	2,065,000
Illinois	25,000	564,000	2,903,000
Pennsylvania	14,000	306,000	1,485,000
Utah	9,000	244,000	1,037,000
Idaho	10,000	227,000	1,008,000
United States	434,000	9,133,000	40,801,000

* Commercial production, including peas grown for seed and cannery peas harvested dry

† Production for processing, including peas for freezing, canning, and other such uses

Wyoming, and Minnesota. Seed for all garden varieties is grown primarily in the same areas that grow the dry commercial peas. Washington and Idaho produce more than 84% of the total seed and dry commercial pea crop. The two main export markets for dry peas are Europe and Latin America. Each year Latin America buys 300,000–350,000 bags weighing 100 lb each. The European market is somewhat sporadic, depending on the total production of Europe's usually large acreage of peas. See LEGUME; VEGETABLE GROWING. [W.A.B.]

Pea diseases. The pea plant is subject to numerous diseases caused by bacteria, fungi, nematodes, and viruses. Some of these diseases can be controlled by seed treatments and planting of resistant varieties; for many, however, no control measures are available.

Root diseases. Root rots in the northern pea canning areas of the United States are caused by *Aphanomyces euteiches*, *Fusarium solani* f. *pisi*, *Pythium* spp., *Ascochyta pinodella*, and *Rhizoctonia solani*; in the southern states root rots are caused by *Sclerotium rolfsii* and *Phymatotrichum omnivorum*. These fungi probably live permanently in the soil and attack many crops other than peas; the root rots therefore constitute a major problem in pea production, control being almost impossible *Aphanomyces euteiches*, the most prevalent and severe of the root rots, can survive in field soils for 10–20 years. Present control measures are limited to keeping field infestation to a minimum by cultural practices and to using a special land-selection program to avoid fields already severely infested.

Two wilt diseases affect peas, common wilt caused by *Fusarium oxysporum* f. *pisi* race 1, and near-wilt caused by *Fusarium oxysporum* f. *pisi* race 2. In both diseases, the organism becomes established in the soil and infects the roots of the pea plant. The organism grows through the water-conducting vessels up into the stem, thus interfering with the passage of water through the stems and into leaves Affected plants wilt, the lower leaves turn yellow, and the plant dies either in the early stage of development or soon after flowering. Wilt diseases are best controlled by the use of resistant varieties. Both market-garden and canning varieties of peas that are resistant to common wilt are available but only a few canning varieties are resistant.

The root knot nematode (*Heterodera marioni*) attacks the roots of peas, causing a reduction in growth and yield and often killing the plants. The disease is most prevalent in the southern and southwestern United States

Foliage diseases Bacterial blight (*Pseudomonas pisi*), anthracnose (*Colletotrichum pisi*), Septoria blotch (*Septoria pisi*), Mycosphaerella blight (*Mycosphaerella pinodes*), and Ascochyta leaf and pod spot (*Ascochyta pisi*) are some of the more important foliage diseases other than the mildews and viruses Most of these pathogens are known to live from one season to the next in and on seed and in the plant refuse left in the field from the previous crop. Partial control is obtained through

the use of clean, treated seed, crop rotation, and

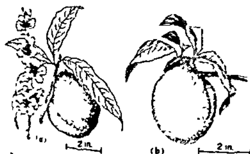
outside the pea-growing areas of the Northwest. There are no effective control measures for downy mildew, but fungicides are sometimes used for powdery mildew control.

The virus diseases of peas may cause severe damage in some seasons. They are generally widespread, the mosaic types being the most prevalent. Varieties resistant to some of the viruses are available. See NEMATODA; PLANT DISEASE; PLANT VIRUS [T.H.K.]

Peach

A native of China, the peach (*Prunus persica*) is adapted to the temperate zone where winter temperature does not go below −15°F. At and below this temperature the wood is killed, especially in the crotches of the trees. Dormant fruit buds are injured by temperatures of 0 to −10°F, and the character of early blossoming makes the peach susceptible to damage from spring frost. In subtropical regions not only is the fruit of poor quality, but winter chilling may not be sufficient to break the dormant period required for some varieties (*see* PLANT GROWTH). For these reasons, commercial culture is confined to the less rigorous parts of the temperate zone, and to areas where nearby bodies of water may exert a modifying influence on climate. In Europe the peach is adapted to Italy, Spain, and southern France. Principal peach-growing areas in North America are (1) central California on the Pacific Coast, with some production in Washington, Oregon, Utah, and Colorado; (2) Maryland, Delaware, and southern New Jersey; (3) Georgia and the Carolinas; (4) Tennessee, Kentucky, and southern Illinois and Indiana; and (5) southern New England, the Niagara region of Canada, western New York, and western Michigan.

Propagation and cultivation. The peach is propagated by budding on peach seedlings produced largely from seed of the Lovell and Muir varieties, and to some degree, on wild forms from the Appalachian Mountains. *See* BUDDING; SEED (BOTANY).



Peaches. (a) Chinese peach grown from seed of wild tree in China (×⅔). (b) Elberta peach (nearly ×½). From L. H. Bailey, ...ndard Cyclopedia of ...iculture.

The peach does best on a well-drained, sandy, or gravelly loam soil, but it also grows on clay loams. Trees are planted 18–22 ft apart. They should be headed (apical bud removed) at about 18–24 in. from the ground and trained to an open or V-shape form with 3–5 scaffold (main lateral) branches. When the tree comes into bearing, the 1-year-old wood (new wood) is commonly shortened back about a half. Since fruit is borne on 1-year-old branches, any cutting back serves not only to develop a compact tree but also to thin the fruit, thus preventing overloading and breakage of the branches. Blossom thinning with poles and brush brooms is an effective thinning practice.

Harvesting. *Fruit is harvested when it loses its* dark green color and begins to show a faint yellowing. Principal American commercial freestone varieties are Dixired, Dixigem, Redhaven, Golden Jubilee, Halehaven, Early Elberta, Elberta, J. H. Hale, and Shippers Late Red. Principal clingstone varieties for canning are Walton, Paloro, Peak, Phillips Cling, Libbee, Stanford, and Ellis.

Commercial peach production in the United States was 62,741,000 bushels in 1957, of which 36,566,000 bushels were produced on the Pacific Coast, including 22,585,000 bushels of clingstone peaches produced in California. Returns to growers were $2.12 a bushel. *See* FRUIT (BOTANY); FRUIT (TREE). [H.B.T.]

Peach diseases. Brown rot, caused by *Monilinia fructicola* and *M. laxa*, is a destructive fungus disease of the peach throughout the world. Blossoms, twigs, and fruit are infected. Major loss is from the decay of the fruit which is converted into a soggy mass unfit for human consumption. Control is achieved by spraying or dusting the trees at regular *intervals with powdered sulfur or with captan, an* organic fungicide (*see* FUNGISTAT AND FUNGICIDE). For best results it is essential also to control the plum curculio, a common fruit insect, whose punctures facilitate entrance of the fungus into the fruit (*see* COLEOPTERA).

In contrast to brown rot, which requires strenuous efforts for its control, peach scab (*Cladosporium carpophilum*), although universally present, is readily controlled by one or two sprays of sulfur shortly after the blossom petals drop. Leaf curl caused by *Taphrina deformans*, a leaf-distorting fungus, can be prevented by one spray of lime sulfur, bordeaux mixture, or ferbam applied before the buds begin to swell.

Bacterial spot, caused by Xanthomonas pruni, is a serious disease of peaches in the United States, China, Japan, and New Zealand. The bacteria kill small groups of cells in the leaves, twigs, and fruit. Infected leaves drop prematurely and this devitalizes the trees mainly through reduced photosynthesis, the most serious long-time effect of the disease. Diseased fruit is edible, but its appearance is marred and its market value reduced. The bacteria survive the winter in small cankers formed on the twigs. Control is difficult because of the prolonged infection period. A mixture of zinc sulfate and

hydrated lime, applied as a spray at intervals of 10–14 days, reduces the severity of the infection in most years.

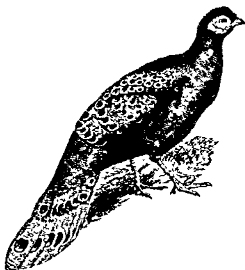
Other fungus diseases affecting peaches are rust (*Tranzschelia discolor*), peach blight (*Coryneum beijerinckii*), and mildew (*Sphaerotheca pannosa*). These diseases cause serious losses in various parts of the world. *Rhizopus nigricans*, occurring throughout the world, is the cause of the rapid decay of harvested and stored peaches.

Among the more than 20 virus diseases known to affect peaches in the United States, yellows, rosette, phony peach, and mosaic are the most serious. The first two kill the trees in a few years, whereas phony peach and mosaic reduce both quality and quantity of fruit and the trees eventually become worthless. *See* PLANT VIRUS.

Peaches are also injuriously affected by nutrient deficiencies in the soil, the symptoms of which often resemble those resulting from disease organisms (*see* PLANT, MINERALS ESSENTIAL TO). Fluctuations in temperature during the winter, particularly in the southern states, frequently result in the death of many peach trees. *See* FRUIT (TREE) DISEASES; PLANT DISEASE CONTROL. [J.C.DU.]

Peacock

A pheasant of the family Phasianidae, *Pavo cristatus*, also called peafowl. The peacock is native to southeastern Asia and Java, but is widely reared in



The peacock, *Pavo cristatus*; length to 6 ft (From P M Duncan, ed , *Cassell's Natural History*, Cassell)

captivity. The adult male is about 6 ft long, 4 ft of which is upper tail-coverts; the true tail is of normal size and is obscured beneath this mass of ornamental feathers Each of the covert feathers has the peacock eye design. Females lack the ornamental tail. Their natural foods are grains, seeds, reptiles, and insects. The white peacock is a mutant variation of this species; the green peacock, *P. muticus*, is a related species, not widely distributed *See* GALLIFORMES. [J.D.B.]

Peanut

A self-pollinated, 1- to 6-seeded legume which is cultivated throughout the tropical and temperate climates of the world (*see* LEGUME). The oil, expressed from the seed, is of high quality, and a large percentage of the 10,000,000-ton annual world production is used for this purpose. In the United States some 65% goes into the cleaned and shelled trade, the end products of which are roasted or salted peanuts, peanut butter, and confections.

Origin and description. Peanuts originated in Bolivia and northeastern Argentina where a large number of wild forms are found. The cultivated species, *Arachis hypogaea*, was grown extensively by Indians in pre-Columbian times. Merchant ships carried seed to many continents during the early part of the sixteenth century. Although grown in Mexico before the discovery of America, the peanut was introduced to the United States from Africa along with the slave trade.

Botanically, peanuts may be divided into three main types: Virginia, Spanish, and Runner, based on branching order and pattern and the number of seeds per pod. *See* SEED (BOTANY).

The peanut's most distinguishing characteristic is its yellow, papilionaceous (resembling a butterfly) flowers which are borne above ground *See* FLOWER (BOTANY) Following fertilization the flower wilts and, after a period of 5–7 days, a positively geotropic (curving earthward) peg or ovary emerges (*see* PLANT MOVEMENTS). After penetrating the soil for a depth of 2–7 cm, the peg assumes a horizontal position and the pod begins to form (Fig. 1).

The pod is a 1-loculed, legume which splits under pressure along a longitudinal ventral suture. *See* FRUIT (BOTANY). Pod size varies from 1×0.5 cm to 2×8 cm, and seed weight varies from ⅕ to 5 grams. The number of seed per pod usually is 2 for the Virginia type, 2–3 for the Spanish, and 3–6 for the Valencia.

The plant may be upright, prostrate, or intermediate between these forms. The main stem is

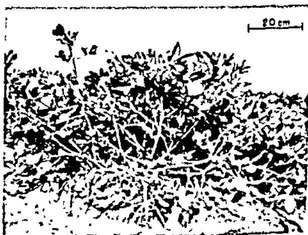


Fig 1. A typical Virginia-type peanut. Note relationship of pod and pegs to plant

usually upright and may be very short in some [illegible]

rangement on stems).

Harvesting and value. Peanuts are harvested by running a special wing type plow under the plants (see AGRICULTURAL MACHINERY). After wilting they are either stacked or allowed to cure in windrows before picking. *See* AGRICULTURAL SOIL AND CROP PRACTICES

The main production areas in the United States extend southward from Virginia and westward to Oklahoma and Texas, with Georgia the largest producer Total annual value of the crop amounts to about $160,000,000. See AGRICULTURAL SCIENCE (PLANT) [A.P.]

Peanut diseases. Yields of peanut hay and fruit are reduced by at least one-fourth because of nonparasitic, insect, bacterial, fungous, nematode, and virus-caused disorders

Nonparasitic Calcium deficiency may initiate fruit decay A deficiency of manganese causes chlorosis and necrotic spots on peanut leaves (Fig. 2) Mechanical injury to the seed radicle causes a curvature of the hypocotyl and retards foliar growth (Fig 2). Radiant energy from the sun causes heat canker on young seedlings. *See* PLANT, MINERALS ESSENTIAL TO.

Insects The southern corn rootworm, *Diabrotica undecimpunctata howardii*, acts as an inoculating agent for the fungi and bacteria that cause fruit rot. The potato leafhopper, *Empoasca fabae*, secretes a toxic substance on the leaves and causes a disorder known as hopperburn. The tobacco thrip,

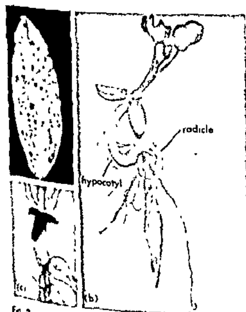


Fig. 2 Nonparasitic diseases of peanut. (a) Necrotic areas caused by manganese deficiency. (b) Multiple primary roots and curled hypocotyl of a seedling with an injured radicle (c) Heat canker. (Photograph by L W Boyle)

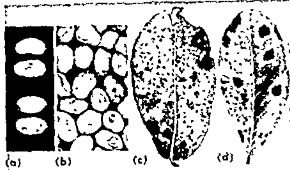


Fig 3. Fungus diseases of peanut. (a) Concealed damage of the cotyledon interfaces (b) Blue damage of the seed coat and cotyledon (photograph by D C. Norton) (c) Rust of leaf (photograph by B B. Higgins). (d) Leafspot caused by *Cercospora arachidicola* (photograph by M. McB Miller)

Frankliniella fusca, causes puckered leaflets and retards the growth of seedlings. *See* INSECTA.

Bacteria Bacterial wilt, caused by *Pseudomonas solanacearum*, occurs in all peanut growing areas and occasionally causes significant losses Several peanut varieties that are resistant to bacterial wilt have been developed in Java

Fungi Species of *Aspergillus, Rhizopus, Mucor, Diplodia, Fusarium, Pythium, Rhizoctonia, Sclerotium, Botrytis*, and *Phymatotrichum* either cause, or play an important role in, the development of rots of planted seed, rots of root, pegs and fruit, a collar rot of the hypocotyl and lower stem, concealed damage of the cotyledon interfaces, and blue damage discoloration of the seed coat and cotyledon of harvested fruit (Fig 3). Rust of the foliar parts, caused by *Puccinia arachidis*, occasionally causes serious plant damage in South America, the West Indies, and southern Texas. The leafspot disease, caused by *Cercospora arachidicola* and *Cercospora personata*, results in premature leaf fall and is the most common and one of the most destructive diseases wherever the crop is grown. Southern blight, caused by *Sclerotium rolfsii*, occurs in all peanut-growing areas, and occasionally causes severe losses This soil borne fungus kills the succulent tissues of the hypocotyl, stem, branches, peg, and fruit *See* STEM (BOTANY).

Nematodes The northern rootknot nematode, *Meloidogyne hapla*, is a common pest of the peanut It feeds inside the root, peg, and fruit and causes small galls on these parts (Fig. 4). The peanut rootknot nematode, *Meloidogyne arenaria*, causes large galls to form on all underground parts. It is known to occur in the United States and Africa and is far more injurious than the northern rootknot nematode. The sting nematodes, *Belonolaimus gracilis* and an unidentified species of *Belonolaimus*, are very destructive; however, they are known to occur only in the lighter soils of the United States. Sting nematodes feed on the outside of the peanut root, peg, and fruit and cause a reduced root system and smaller fruit. The smooth-

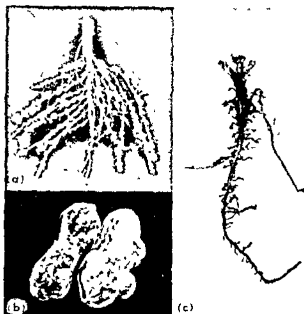


Fig. 4. Nematode diseases of peanut. (a) Root infected by the peanut rootknot nematode. (b) Fruit infected by the peanut rootknot nematode. (c) Root injury caused by the feeding of the sting nematode.

headed meadow nematode, *Pratylenchus brachyurus*, feeds inside the peanut root, peg, and fruit and causes necrotic lesions on these parts.

Viruses. Viruses causing either rosette, mosaic, ringspot, or stunt of peanut have been reported from all the principal peanut-producing areas. However, no thorough study has been made of these diseases, nor are their effects on yield well known, except that of rosette which causes severe losses in Africa. See PLANT DISEASE; PLANT DISEASE CONTROL; PLANT VIRUS [L I M]

Processing. Peanut processing begins in the fields where the peanuts are removed from vines by portable, mechanical pickers. The freshly dug peanut plants are windrowed in the field, and the peanuts, with a moisture content of 25–30%, are picked from the vines within 3 days.

The total quality of edible peanuts processed in the United States annually from 1938 to 1956 averaged 1,475,000,000 lb Of this 41% was the Spanish variety, 30% Runner and 29% Virginia At the same time an average of 290,000,000 lb was crushed for oil, with a production of 112,000,000 lb of crude oil

The 1946–1955 average production of peanuts by states in millions of lb was Georgia, 586; Alabama, 245; Texas, 244; North Carolina, 276, Virginia, 209; and other states, 193 With few exceptions processing is near centers of production.

Both edible and oil stock peanuts are processed the year around. The 1951–1957 average for edible peanuts, reached a peak of 155,000,000 lb in November, and a low of 45,000,000 lb in August. Crushing for oil over the same period reached a peak of 18,000,000 lb in February, and a low of 7,000,000 lb in October.

The average 1955–1957 distribution of edible peanuts by products was peanut butter, 51.2%; salted peanuts, 25.1%; peanut candy, 21 8%; and other products, 1.9%. The composition of peanuts of all varieties is 44–48% oil, 25–30% protein, 5–7% water, 2–4% carbohydrates, a good supply of phosphorus, calcium, and niacin, and a trace of iron, thiamine, and riboflavin.

Cleaning. Peanuts from the pickers are delivered to warehouses for cleaning This consists of removing sticks, stems, small rocks and faulty nuts by a series of screens and blowers. The operation reduces the bulkiness of the nuts by 10–20%.

Storing. Cleaned peanuts are stored unshelled, in silos or warehouses, for continuous shelling and delivery to end-users; and shelled, in refrigerated warehouses, at 32–36°F with 65% relative humidity (r.h.). Refrigeration insures protection against insects and rancidity.

Shelling. This consists of breaking the shells by passing the nuts between series of rollers. The shells and small, immature pegs are separated by screens and blowers, and the discolored kernels are removed by electric eyes. Shelling reduces the weight of peanuts 30–60%, the space occupied 60–70%, and the shelf-life 60–75%, depending upon the variety.

Blanching. This consists of removing the seed coats and usually the hearts of peanuts. There are [illegible]

[illegible] lowed by rubbing the nuts between soft surfaces and removing the skins by blowers and the hearts by screens.

Water blanching consists of arranging the nuts longitudinally in troughs, passing them beneath spring fingers with blades, and slitting the skins from end to end. The skins are removed as a spiral conveyor carries the nuts through a 1-minute bath of scalding water. The nuts are dried to 7% moisture prior to storage or manufacture into peanut products

Dry roasting. Peanuts for use in peanut butter or bakery products are dry-roasted to develop desirable color, texture, and flavor Unblanched peanuts are heated to 204 4°C. for 20–30 minutes, after which they are cooled and blanched

Peanut butter Shelled, ground, parched peanuts were first prepared about 1890 as food for infants and invalids. From a kitchen operation this has become a major industry, with individual plants manufacturing 10,000,000 lb of peanut butter annually (Fig. 5) The product consists of blanched, dry-roasted peanuts, ground to a size to pass through a 200-mesh screen. Additives to improve smoothness, spreadability, and flavor include 1½% salt, 0.1¼% hydrogenated vegetable oil, 2% dextrose, 2–4% corn syrup or honey. Additives to improve nutritive qualities include 185 ml/100 grams ascorbic acid and yeast.

Shelled peanuts are dry-roasted in a gas-fired, rotary roaster at 400°F. When cooking is complete, the peanuts are dumped into a blower-cooler vat where they are brought to 85°. The cool peanuts

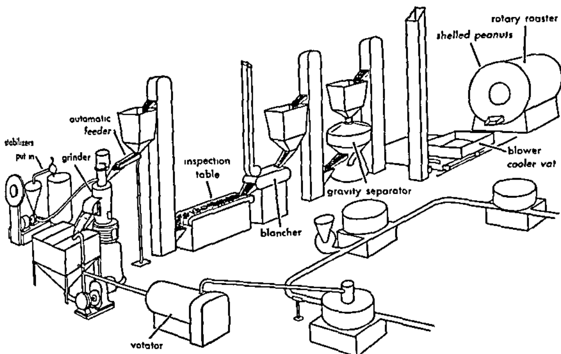


Fig 5 Manufacture of peanut butter (From *Food Processing*)

pass through a gravity separator which removes foreign material; and to a blancher which removes the hearts by a shaker-screen and the skins by a blower After passing over the inspection table the nuts pass through an automatic feeder and into a grinder Stabilizers are metered into the mill simultaneously with the peanuts, and are thoroughly dispersed in the butter. The stabilized peanut butter is cooled in a votator (rotating refrigerated cylinder) then automatically packed, labeled and stored

Peanut butter contains 50–52% fat, 28–29% protein 2–5% carbohydrate, and 1–2% moisture. *See* ASCORBIC ACID, YEAST.

Oil roasting Peanuts for salting are roasted in coconut oil or partly hydrogenated vegetable oil at 148 9°C for 15–18 minutes The end-point is based on change of color and is controlled electrically or manually

Salting Peanuts either blanched or unblanched are roasted in oil and salted Finely ground salt (2-3%), and an oil base binder, is mixed with freshly cooked nuts, which are then placed in flexible bags or canned under vacuum.

Salting in the shell. Peanuts may be salted in the shell. This involves soaking in a surface-active agent at 60°C for 15 minutes, rinsing, submerging in saturated brine, and subjecting nuts to 20 in. vacuum two or three times for 5 minute periods, rinsing and drying

Extraction of oil The recovery of oil from peanuts is by either of three methods—hydraulic pressing, expeller pressing, or solvent extraction Hydraulic pressing is essentially the same as used for oil recovery from cottonseed The peanuts are broken between rollers, and the shells are removed by blowers The meats are crushed and heated under 25 lb steam pressure for 10 minutes, stabilized at 7% moisture, and pressed at 137.8°C. The yield of oil by hydraulic pressing is 41–47%, and the press cake contains 42–45% protein, 5–6% moisture, and 7–8% oil

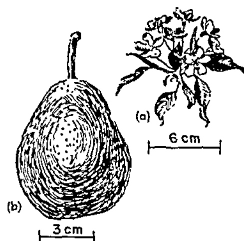
Solvent extraction of oil from ground peanuts is similar to that from soybeans (*see* SOYBEAN). The direct-solvent process was introduced in the late 1940s, and was followed in 1950 by the prepress-solvent process, which in turn was followed in 1954 by the high-speed screw press The yield of oil by solvent extraction is 48–50%, and the meal contains 50–52% protein, 1–2% oil and 1–2% moisture *See* FOOD ENGINEERING [J.G.W]

Bibliography: K H. Garren and C Wilson, *The Peanut, The Unpredictable Legume*, 1951.

Pear

A fruit native to the region from the Caspian Sea westward into Europe, and of very old culture, being known nearly 1000 years before the Christian era It spread across Europe and was extremely popular in Belgium and France during the eighteenth and nineteenth centuries Early settlers in North America made extensive efforts to grow pears The Pacific Coast was eventually proven to be one of the best habitats.

Commercial types. There are four commercial types of pears (1) *Pyrus communis*, the European pear, including all the old standard varieties; (2) *Pyrus serotina*, the Asian or Oriental sand pear, with characteristic roundish shape and long-keeping, gritty, poor-quality flesh, (3) hybrids between these two species, represented by Keiffer and Leconte; and (4) *Pyrus nivalus*, the snow pear,



Pear, *Pyrus communis.* (*a*) Cluster of flowers. (*b*) The pyriform (pear-shaped) fruit of the Bartlett pear. (*From L. H. Bailey, ed., The Standard Cyclopedia of Horticulture, Macmillan, 1937*)

which is grown in Europe for cider or perry (a fermented liquor).

Propagation and cultivation. The pear is propagated by budding onto seedlings grown from seeds of wild pears from Europe (French pear) and from such cultivated varieties as Bartlett and Winter Nelis (*see* BUDDING). Several Oriental rootstocks have been tried because of their resistance to the dread fire-blight disease, *Bacillus amylovorus*, but the fruits borne by trees propagated on these rootstocks are afflicted with a trouble known as black end, which makes them unmarketable. The pear may be dwarfed very successfully by propagating on quince roots, *Cydonia oblonga. See* GRAFTING OF PLANTS; QUINCE.

The pear does best in an equable climate with sufficient summer heat to develop good quality. Trees are injured by temperatures below −10 to −15°F, but require winter cold to break the dormant period (*see* PLANT GROWTH). Commercial production is, accordingly, confined to temperate-zone regions where winter cold is not too severe, as in California and the Pacific Northwest, or where large bodies of water temper the climate, as in western New York, western Michigan, and the Niagara Peninsula of Canada.

Harvesting and commercial production. For best quality, pears are picked early and ripened off the tree. Such winter varieties as Doyenne du Comice, Passe Crassane, Beurre Hardy, Beurre Bosc, and Beurre d'Anjou are held at 32°F, but must thereafter be ripened for several days at 60°F to develop high quality.

Commercial production in the United States was 32,005,000 bushels in 1957, of which 29,047,000 bushels were produced on the Pacific Coast. Returns to growers were $2.01 per bushel. The Bartlett pear constituted 21,347,000 bushels of this production. Other important varieties are Doyenne du Comice, Beurre Hardy, Beurre Bosc, Beurre d'Anjou Kieffer, and Seckel. The Williams Bon Chretien variety of Europe is the same as the Bartlett. *See* FRUIT (BOTANY); FRUIT (TREE); FRUIT (TREE) DISEASES; PLANT DISEASE CONTROL. [H B.T.]

Pearl

The term pearl is applied properly to any mollusk-formed calcareous concretion that displays an orient and is lustrous. There are two major groups of bivalved mollusks in which gem pearls may form: the salt-water pearl oyster, *Pinctada*, of which there are several species; and a number of genera of fresh-water clams. Usually, jewelers refer to salt-water pearls as Oriental pearls, regardless of their place of discovery, and to those from fresh-water bivalves as fresh-water pearls. *See* GEM.

Formation. Between the body mass and the valves of the mollusk extends a curtainlike tissue called the mantle. Epithelial cells on the side of the mantle toward the shell perform the several stages of the shell-secreting process during the life of the mollusk. One of the stages of shell building is the secretion of nacre, the colorful, lustrous, mother-of-pearl material. In order for a pearl to form, a tiny object such as a parasite or a grain of sand must work through the mantle, carrying with it epithelial cells. When this happens, secretion of nacre around the invading object builds a pearl within the body of the mollusk. Whole pearls form within the body mass of the mollusk, in contrast to blister pearls, which form as protrusions on the inner surface of the shell. Edible oysters produce lusterless concretions, but never pearls.

Pearls occur in a great variety of shapes. The term baroque is used for the common, irregularly shaped forms. The most common and most desirable shape is the spherical or nearly spherical; this is the shape usually chosen for necklaces. Other desirable shapes include those called button, pear, egg, and drop. Particularly desirable colors include cream, rose, white, black, and gold.

Pearls are composed of many tiny overlapping plates of nacreous material. Nacre consists of prismatic pseudohexagonal aragonite crystals (oriente so that the long crystallographic axis is at rigl angles to the plane of the platelet) held together l conchiolin, a hornlike organic secretion. Chemica analyses of pearls show calcium carbonate, organi material, and water; the relative quantities var with the species of the mollusk, the position of th pearl within the shell, and other factors. CaCC content is usually from 90 to 92%, but may be som what lower. Organic matter usually makes up fror 4 to 6% and water from 2 to 4%. *See* ARAGONITE.

Producing areas. The major pearl-producing re gion in the world today is the Persian Gulf. Fron its pearl fisheries come most of the natural pearl used for gem purposes, but recent output is only a fraction of that of a long period during the nine teenth century through the 1920s. Since World War II, the sale of Oriental pearls has never re gained earlier peaks. In the Persian Gulf, the im portant species *Pinctada margaritifera* is found at depths of from about 4 to 8 fathoms on broad banks that extend for many miles into the Gulf from

both shores Since pearl-producing mollusks are only one of many types of mollusk, fishing is a great gamble, and the waste of nonbearing mollusks is enormous. In times of demand, depletion of the mollusk supply is a grave problem The supply has apparently never fully recovered from the depletion caused by the heavy fishing of the 1920s Less important pearl sources include the coast of northern Venezuela, to the north of Australia, between Ceylon and India, in the Red Sea, and in the South Pacific.

Many of the rivers of the central portion of the United States have pearl-bearing mussels. In the nineteenth century, several rich finds were made that led to rapid exploitation and virtual exhaustion in those regions of the supplies of pearl mollusks; these were principally of the genus *Unio* Fresh-water pearls vary in color from almost pure white to many that are more strongly tinted than the majority of the salt-water variety

Cultured pearl. The substitute for natural pearls, to which the name cultured pearl has been given, is usually made by inserting a large bead into a mollusk to be coated with nacre When the pearl-bearing mollusk, *Pinctada martensii*, reaches maturity at three and one-half years, the mollusks are gathered and prepared for pearl cultivation. Workers trained for the task cut a channel into the foot mass and insert a large sphere (bead), plus a small section of mantle tissue, with the epithelial cells next to the bead Beads are prepared from large American fresh-water shells The mollusks are placed in cages suspended from rafts in sheltered bays and usually left for three and one-half to four years, except for periodic cleaning and inspection The rate of nacre accretion in Japanese waters is only about 0.15 millimeter annually, so that the diameter increases [illegible]

[illegible]

[illegible] Baroque fresh water pearls have been produced without bead nuclei in Japan. See PELECYPODA [R.T L.]

Peat

A dark brown or black residuum produced by the partial decomposition and disintegration of mosses sedges, trees, and other plants that grow in marshes and other wet places Forest type peat, when buried and subjected to geological influences of pressure and heat, is the natural forerunner of most coal Peat may accumulate in depressions such as the coastal and tidal swamps in the Atlantic and Gulf Coast states, in abandoned ox bow lakes where sediments transported from a distance are deposited, and in depressions of glacial origin Moor peat is formed in relatively elevated, poorly drained moss-covered areas as in parts of northern Europe. See COAL; HUMUS.

In the United States, where the principal use of peat is for soil improvement, the estimated reserve on an air-dried basis is 13 827,000 short tons. In Ireland and Sweden peat is used for domestic and even industrial fuel. In Germany, peat is the source of low-grade montan wax. [C H.C.]

Pebble mill

A tumbling mill that grinds or pulverizes materials without contaminating them with iron (*see* TUMBLING MILL). Because the pebbles have lower specific gravity than steel balls, the capacity of a given size shell with pebbles is considerably lower than with steel balls The lower capacity results in lower power consumption. The shell has a nonmetallic lining to further prevent iron contamination, as in pulverizing ceramics or pigments. Selected hard pieces of the material being ground can be used as pebbles to further prevent contamination [R.M H.]

Pecan

A large deciduous tree (*Carya illinoensis*), native to North America, and its fruit, a true nut. In the United States large-scale commercial plantings have been made in Florida, Georgia, Alabama, Mississippi, Louisiana, and Texas, with smaller acreages in New Mexico, Arizona, and California.

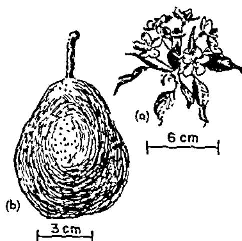
There are two distinct botanical types, one native to the states bordering the Gulf of Mexico and to northern and central Mexico, and the other to the Ohio River valley. The southern pecan must have long hot summers for proper maturity of kernels Northern pecans require less summer heat and will withstand colder winters, but the nuts are generally smaller and have thicker shells Some intermediate types are found in Texas and Oklahoma. A large percentage of the pecans produced in the world are grown in the United States, the southern states being the main source of supply. A few are also grown in Mediterranean countries, Australia, and South Africa.

Pecans have a high oil content and a pleasing flavor They are consumed in large volume in the United States as dried nuts, raw or roasted, in candies and cookies, as toppings for bakery goods, in ice cream, and as a meat substitute.

Pecan trees grow up to 200 ft high, with a crown spread of 100 ft under favorable conditions In



Pecan (a) Twig with leaves and fruit. (b) Hulled nuts.



Pear, Pyrus communis. (a) Cluster of flowers. (b) The pyriform (pear-shaped) fruit of the Bartlett pear. (From L. H. Bailey, ed., The Standard Cyclopedia of Horticulture, Macmillan, 1937)

which is grown in Europe for cider or perry (a fermented liquor).

Propagation and cultivation. The pear is propagated by budding onto seedlings grown from seeds of wild pears from Europe (French pear) and from such cultivated varieties as Bartlett and Winter Nelis (*see* BUDDING). Several Oriental rootstocks have been tried because of their resistance to the dread fire-blight disease, *Bacillus amylovorus*, but the fruits borne by trees propagated on these rootstocks are afflicted with a trouble known as black end, which makes them unmarketable. The pear may be dwarfed very successfully by propagating on quince roots, *Cydonia oblonga*. *See* GRAFTING OF PLANTS; QUINCE.

The pear does best in an equable climate with sufficient summer heat to develop good quality. Trees are injured by temperatures below −10 to −15°F, but require winter cold to break the dormant period (*see* PLANT GROWTH). Commercial production is, accordingly, confined to temperate-zone regions where winter cold is not too severe, as in California and the Pacific Northwest, or where large bodies of water temper the climate, as in western New York, western Michigan, and the Niagara Peninsula of Canada.

Harvesting and commercial production. For best quality, pears are picked early and ripened off the tree. Such winter varieties as Doyenne du Comice, Passe Crassane, Beurre Hardy, Beurre Bosc, and Beurre d'Anjou are held at 32°F, but must thereafter be ripened for several days at 60°F to develop high quality.

Commercial production in the United States was 32,005,000 bushels in 1957, of which 29,047,000 bushels were produced on the Pacific Coast. Returns to growers were $2.01 per bushel. The Bartlett pear constituted 21,347,000 bushels of this production. Other important varieties are Doyenne du Comice, Beurre Hardy, Beurre Bosc, Beurre d'Anjou Kieffer, and Seckel. The Williams Bon Chretien variety of Europe is the same as the Bartlett. *See* FRUIT (BOTANY); FRUIT (TREE); FRUIT (TREE) DISEASES; PLANT DISEASE CONTROL. [H.B.T.]

Pearl

The term pearl is applied properly to any mollusk-formed calcareous concretion that displays an orient and is lustrous. There are two major groups of bivalved mollusks in which gem pearls may form: the salt-water pearl oyster, *Pinctada*, of which there are several species; and a number of genera of fresh-water clams. Usually, jewelers refer to salt-water pearls as Oriental pearls, regardless of their place of discovery, and to those from fresh-water bivalves as fresh-water pearls. *See* GEM

Formation. Between the body mass and the valves of the mollusk extends a curtainlike tissue called the mantle. Epithelial cells on the side of the mantle toward the shell perform the several stages of the shell-secreting process during the life of the mollusk. One of the stages of shell building is the secretion of nacre, the colorful, lustrous, mother-of-pearl material. In order for a pearl to form, a tiny object such as a parasite or a grain of sand must work through the mantle, carrying with it epithelial cells. When this happens, secretion of nacre around the invading object builds a pearl within the body of the mollusk. Whole pearls form within the body mass of the mollusk, in contrast to blister pearls, which form as protrusions on the inner surface of the shell Edible oysters produce lusterless concretions, but never pearls.

Pearls occur in a great variety of shapes The term baroque is used for the common, irregularly shaped forms. The most common and most desirable shape is the spherical or nearly spherical; this is the shape usually chosen for necklaces Other desirable shapes include those called button, pear, egg, and drop. Particularly desirable colors include cream, rose, white, black, and gold

Pearls are composed of many tiny overlapping plates of nacreous material Nacre consists of prismatic pseudohexagonal aragonite crystals (oriented so that the long crystallographic axis is at right angles to the plane of the platelet) held together by conchiolin, a hornlike organic secretion. Chemical analyses of pearls show calcium carbonate, organic material, and water; the relative quantities vary with the species of the mollusk, the position of the pearl within the shell, and other factors. $CaCO_3$ content is usually from 90 to 92%, but may be somewhat lower Organic matter usually makes up from 4 to 6% and water from 2 to 4%. *See* ARAGONITE.

Producing areas. The major pearl-producing region in the world today is the Persian Gulf. From its pearl fisheries come most of the natural pearls used for gem purposes, but recent output is only a fraction of that of a long period during the nineteenth century through the 1920s. Since World War II, the sale of Oriental pearls has never regained earlier peaks. In the Persian Gulf, the important species *Pinctada margaritifera* is found at depths of from about 4 to 8 fathoms on broad banks that extend for many miles into the Gulf from

both shores Since pearl-producing mollusks are only one of many types of mollusk, fishing is a great gamble, and the waste of nonbearing mollusks is enormous In times of demand, depletion of the

important pearl sources include the coast of northern Venezuela, to the north of Australia, between Ceylon and India, in the Red Sea, and in the South Pacific.

Many of the rivers of the central portion of the United States have pearl-bearing mussels In the nineteenth century, several rich finds were made that led to rapid exploitation and virtual exhaustion in those regions of the supplies of pearl mollusks; these were principally of the genus *Unio* Fresh water pearls vary in color from almost pure white to many that are more strongly tinted than the majority of the salt-water variety.

Cultured pearl. The substitute for natural pearls, to which the name cultured pearl has been given, is usually made by inserting a large bead into a mollusk to be coated with nacre. When the pearl bearing mollusk, *Pinctada martensii*, reaches maturity at three and one half years, the mollusks are gathered and prepared for pearl cultivation. Workers trained for the task cut a channel into the foot mass and insert a large sphere (bead), plus a small section of mantle tissue, with the epithelial cells next to the bead Beads are prepared from large American fresh-water shells The mollusks are placed in cages suspended from rafts in sheltered bays and usually left for three and one-half to four years, except for periodic cleaning and inspection The rate of nacre accretion in Japanese waters is only about 0.15 millimeter annually, so that the diameter increases about 0 30 mm annually; thus, a 7.2-mm cultured pearl

genus *Pinctada*. Baroque fresh water pearls have been produced without bead nuclei in Japan See PELECYPODA. [R T L]

Peat

A dark brown or black residuum produced by the partial decomposition and disintegration of mosses, sedges, trees, and other plants that grow in marshes and other wet places. Forest-type peat, when buried and subjected to geological influences of pressure and heat, is the natural forerunner of most coal. Peat may accumulate in depressions such as the coastal and tidal swamps in the Atlantic and Gulf Coast states, in abandoned ox-bow lakes where sediments transported from a distance are deposited, and in depressions of glacial origin Moor peat is formed in relatively elevated, poorly drained moss-covered areas as in parts of northern Europe. See COAL; HUMUS

In the United States, where the principal use of peat is for soil improvement, the estimated reserve on an air-dried basis is 13,827,000 short tons In Ireland and Sweden peat is used for domestic and even industrial fuel In Germany, peat is the source of low-grade montan wax [C H.C.]

Pebble mill

A tumbling mill that grinds or pulverizes materials without contaminating them with iron (*see* TUMBLING MILL). Because the pebbles have lower specific gravity than steel balls, the capacity of a given size shell with pebbles is considerably lower than with steel balls. The lower capacity results in lower power consumption. The shell has a nonmetallic lining to further prevent iron contamination, as in pulverizing ceramics or pigments. Selected hard pieces of the material being ground can be used as pebbles to further prevent contamination. [R.M.H]

Pecan

A large deciduous tree (*Carya illinoensis*), native to North America, and its fruit, a true nut In the United States large-scale commercial plantings have been made in Florida, Georgia, Alabama, Mississippi, Louisiana, and Texas, with smaller acreages in New Mexico, Arizona, and California.

There are two distinct botanical types, one native to the states bordering the Gulf of Mexico and to northern and central Mexico, and the other to the Ohio River valley. The southern pecan must have long hot summers for proper maturity of kernels. Northern pecans require less summer heat and will withstand colder

grown in the United States, the southern states being the main source of supply A few are also grown in Mediterranean countries, Australia, and South Africa.

Pecans have a high oil content and a pleasing flavor They are consumed in large volume in the United States as dried nuts, raw or roasted, in candies and cookies, as toppings for bakery goods, in ice cream, and as a meat substitute.

Pecan trees grow up to 200 ft high, with a crown spread of 100 ft under favorable conditions. In

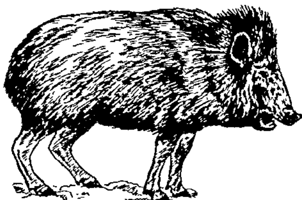


Pecan. (a) Twig with leaves and fruit (b) Hulled nuts

nature they usually grow close to rivers or streams in deep, open-type soils. They withstand flooding well and are not easily blown over because they have a deep, sturdy root system. *See* NUT CROP CULTURE. [E.F.S.]

Peccary

The wild pig or javelina of America. There are two living species in the genus *Tayassu*, family Tayassuidae, in addition to several extinct forms. One species, the collared peccary, *T. tajacu*, occurs in central Texas, south central Arizona, and southward into Patagonia. They are distinctly piglike in



The collared peccary, *Tayassu tajacu*; length 38 in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

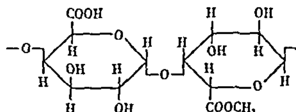
appearance, with short, bristly, grizzled gray and black hair, lighter over the front of the shoulder. They have three toes on each hind foot (pigs have four), and only a vestigial tail. Peccaries are good sport animals and are also of some value for their flesh and hide. *See* ARTIODACTYLA. [J.D.B.]

Pectin

A distinct group of polysaccharides which occur in the cell walls and intercellular layers of all land plants. They are extractable by hot water, dilute acid, ammonium oxalate solutions, and other reagents, are precipitated by alcohol, and are especially noted for their ability to form voluminous gels.

Uses. The most extensive use of pectin is in preparation of gels for different food uses. Commercial designation of pectin as jelly grade refers to the weight of sucrose which one unit weight of pectin at suitable conditions of acidity will form into a jelly containing 65% sugar solids. Good commercial grade pectins have jelly grades between 150 and 300. Extensive cooling is not necessary to gel a high-ester pectin, whereas low-ester pectins will gel when cooled or when multivalent cations are added in small amounts at sugar concentrations much below 65%. Addition of calcium or alum to fresh produce such as pickles causes a firming action attributed to the formation of rigid gels in the outer tissue of the substance treated. Pectins are used to inhibit weeping in thawed frozen fruits. Medical uses have been as an intestinal tract regulator and for intravenous treatment of shock. D-Galacturonic acid is prepared from pectin which in turn can be used to synthesize ascorbic acid (vitamin C). Further uses of pectins are as clarification, thickening, foam-forming, and sizing agents, and for fatty-acid production by suitable fermentation techniques. Plasticized films of pectinic acids are flexible and reasonably strong and can be used as oil-repellent containers or cratings.

Preparation and properties. Pectins from different sources have somewhat different compositions, and some contain a few acetyl groups (beet pectins and some fruit pectins). The major component of all pectins is a polymer of D-galacturonic methyl ester, the methyl ester of a galacturonan. Smaller amounts of a polymer of L-arabinose (an araban) and a polymer of D-galactose (a galactan) are found, also. Galacturonic acid units are not fully esterified. Thus, where the theoretical maximum methoxyl content of the polymeric galacturonic acid, pectic acid, is 16.35%, the methoxyl content of high-ester pectins is less, but still above 8%. Low-ester pectins contain less than 7% methoxyl, usually 3–5%. Immature plant tissues contain water-insoluble pectin, termed protopectin, and as the tissue matures, the pectin becomes more soluble. The gel-forming properties of pectin are related to the presence of galacturonan methyl ester, actually



the partial methyl ester of pectic acid, which is a linear molecule that has a molecular weight of 30,000–300,000 and in which the D-galacturonic acid units are joined by α-($1 \rightarrow 4$) linkages. In the accompanying araban, L-arabofuranose units are joined by α-($1 \rightarrow 5$) and α-($1 \rightarrow 3$) linkages, and in the galactan, D-galactopyranose units are joined by β-($1 \rightarrow 4$) linkages. The linear nature of pectin has been shown by examination of its physical behavior. Tough, pliable films can be prepared. X-ray analysis of fibers prepared from a pectin solution show oriented crystallite patterns. Viscosity measurements, sedimentation, diffusion, and flow birefringence also support this view.

Methyl groups are removed slowly by acid hydrolysis to leave at intermediate stages a partially methylated polysaccharide, pectinic acid, wherein the ester groups are randomly distributed. With the enzyme, pectin esterase, obtained from such sources as roots, leaves, and fruits of all higher plants and also from a number of microorganisms, the ester groups are quickly removed. However, at intermediate stages the molecule contains both segments that have the full complement of ester groups and segments that are completely devoid of ester groups. Random deesterification by alkali is ac-

ompanied by depolymerization and is little used. Pectic acid has low water solubility. Because all pectins have some free carboxyl groups, they are affected by multivalent cations such as calcium ion which can cross-link in salt formation between

must be demineralized to remove such cations When mixed with sugar, high-ester pectins will form gels or jellies, and this property has found commercial value in preparation of jams and marmalades In the United States, most pectin is prepared from apple pomace, citrus waste, or beet pulp After treating the pulp in sulfurous acid to destroy pectic enzymes in the fresh tissue, the material is washed with demineralized water (free of calcium and magnesium ions) and extracted with hot dilute acid (pH 1 0–3 5) for about 30 min. Pectin extracts are usually clarified by treatment with amylolytic enzymes and decolorized with carbon before precipitation into alcohol or metal salt solutions, such as alum If metal salt solutions are used, they must be removed by washing with acidic alcohol Drying of the solid material followed by pulverizing yields a commercially suitable product. Recent use of calcium complexing agents such as

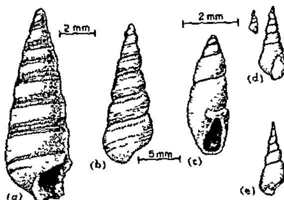
year. world production may be about 50% higher.

Water solutions of purified pectin of up to 2–3% concentration are easily prepared. In solution, these polysaccharide molecules behave as typical colloids The free acid groups of pectin can be titrated directly with dilute alkali solution to give titration curves that resemble those of monobasic acids Dehydrating agents easily precipitate pectins because they are composed of large particles which are essentially gel fragments Precipitation can be effected in certain cases by the presence of other more hydrophilic colloids Pectins are usually characterized by uronic acid content, high viscosity, and gel forming ability. *See* GEL; GUM; POLYSACCHARIDE. [K.W.KI ; R L WH.]

Pectinibranchia

An order of gastropods, also called Mesogastropoda, which contains many important families of snails. Respiration is by means of ctenidia which are composed of several gill leaves in the mantle cavity The nervous system is not concentrated, an operculum is usually present, and the sexes are separate Various families are found in the sea, in fresh water, and on land.

The family Littorinidae is of wide distribution, and the various species live at the high tide line, or, more usually, in the intertidal zone. The common periwinkle, *Littorina littorea L.*, is a species in this family Another important family, the Naticidae, contains the moon shells, with several species of economic importance because they are predatory on edible species of clams. The family Cypraeidae contains the cowries, many species of which are among the most beautifully colored of all marine snails Several important families occur in fresh water, such as Viviparidae, Pilidae, Pleuroceridae, and Thiaridae, whereas other families occur only on the land, for example, the Cyclophoridae and Pomatiasidae. *See* GASTROPODA, PROSOBRANCHIA. [W.J.C.]



Order Mesogastropoda (a) Orthonema (Penn.-Perm). (b) Acanthonema (Dev.). (c) Girtyspira (Miss-Penn). (d, e) Meekospira (Ord.-Perm) (R. R Shrock and W. H. Twenhofel, Principles of Invertebrate Paleontology, 2d ed, McGraw-Hill, 1953)

Pectolite

A mineral inosilicate with composition $Ca_2NaSi_3O_8(OH)$, crystallizing in the triclinic system Crystals are usually acicular in radiating aggregates There is perfect cleavage parallel to the front and basal pinacoids yielding splintery fragments elongated on the *b* crystal axis. The hardness is 5 on Mohs scale, and the specific gravity is 2.75 The mineral is colorless, white, or gray with a vitreous to silky luster. Pectolite is a secondary mineral occurring characteristically in cavities in basalt and associated with zeolites, prehnite, apophyllite, and calcite The most notable localities in the United States are at Paterson, Bergen Hill, and Great Notch, New Jersey *See* SILICATE MINERALS [C S.HU]

Pedipalpi

Formerly an order of the Arachnida which contained the principal types of whip scorpions These animals are now placed in the separate orders Uropygi and Amblypygi *See* AMBLYPYGI; UROPYGI [W J CH]

Pedology

Defined narrowly, a science that is concerned with the nature and arrangement of horizons in soil profiles; the physical constitution and chemical composition of soils; the occurrence of soils in relation to one another and to other elements of the environment such as climate, natural vegetation, topography, and rocks, and the modes of origin of soils Pedology so defined does not include soil technol-

Defined more broadly, pedology is a science that treats soils, their nature, properties, formation, distribution, functioning, and response to use, management, and manipulation.

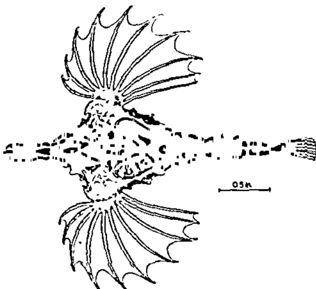
The first definition is used widely in the United States, and less widely in other countries. The second definition is used widely in both the United States and other countries. *See* SOIL; SOIL CONSERVATION; SOIL MECHANICS. [R.W.SI.]

Peening

A metal-finishing operation, also called shot peening, in which small steel shot is thrown against a piece such as a cutting tool The impact of the shot on the work plastically deforms the surface to a depth of a few thousandths of an inch, producing residual compressive stress. The material is thus made more resistant to fatigue failure. Surface hardness of the material is also increased slightly by the cold working produced by the shot. The shot is hurled at high velocity upon the work by centrifugal force or by an air blast. *See* MACHINING OPERATIONS [A T.]

Pegasiformes

The sea moths or sea dragons. The precise relationships of this order of peculiar actinopterygian fishes are unknown. The group is also known as the Hypostomides The body is encased in a broad, bony framework anteriorly and has bony rings posteriorly. Enlarged nasal bones form a rostrum that overlies the small, toothless mouth. The preopercle, pterosphenoid, opisthotic, entopterygoid, and metapterygoid are absent, and the gill cover has a single bone. The greatly expanded horizontal pectoral fin belies its appearance and does not function in aerial gliding. The pelvic fin is abdominal and consists of a spine and two long rays The short, opposed dorsal and anal fins have no spines There is no swimbladder.



Sea moth, *Pegasus draconis*. (After D S. Jordan and J. O. Snyder, vol. 24, Leland Stanford University Contributions to Biology, 1901)

There is a single family, Pegasidae, with one genus, *Pegasus*, and four or five species. They live among vegetation on Indo-Pacific shores from East Africa to Japan and Australia. They rarely exceed 4 in. in length. *See* ACTINOPTERYGII. [R.M.B.]

Pegasus

The Winged Horse, in astronomy, is a northern winter constellation. Pegasus is usually identified by the four bright stars situated on the corners of a large square known as the Great Square in Pegasus. The constellation is represented by a winged horse. Markab (the Saddle), a navigational star, occupies the southwestern corner of the square. The star Alpheratz at the opposite corner is really in the constellation Andromeda. The star at the northwestern corner is a red star, known as Scheat, a giant irregular variable. Diagonally opposite on the southeastern corner of the square is Algenib. Enif, another navigational star, lies in the nose of the horse. *See* CONSTELLATION. [C.S.Y.]

Pegmatite

Generally any extremely coarse-grained, crystalline rock or any body composed of such rock. Pegmatites are relatively small. They range widely in composition and commonly carry numerous rare minerals. They are relatively light-colored rocks, and most are of granitic composition (quartz plus feldspar), corresponding mineralogically to granite, granodiorite, or quartz diorite Pegmatites are principal sources for feldspar, mica, gemstones, and rare elements. Granitic pegmatites may be intimately associated with aplite. *See* APLITE.

Types. Gabbro pegmatite is widespread but not abundant. It forms small pods, pipes, lenses, sheets, or veins enclosed by large masses of diabase, gabbro, or related rocks.

Syenite and nepheline syenite pegmatites are comparatively rare Well known are those of the Kola Peninsula, U.S.S.R., which are rich in the rare earths, zirconium and titanium Similar pegmatites in Norway and Sweden are noted for their wide variety of minerals; many contain lithium, rubidium, cesium, cerium, lanthanum, arsenic, antimony, zirconium, uranium, and thorium.

Granitic pegmatites occur most abundantly as lenticular, tabular, or irregular bodies in metamorphic rocks (schists and gneiss) and less commonly within or slightly marginal to granitic bodies. Pegmatite in granite appears most abundant in rocks of Precambrian age

Shape. Pegmatite bodies range up to several thousand feet long and several hundred feet wide. Boundaries may be smooth or highly irregular, and contacts with adjacent rocks may be sharp or gradational Where formed in strongly foliated or layered rocks, pegmatities are commonly elongate parallel to the direction of layering

Composition. The mineralogical composition is chiefly microcline, microcline perthite, quartz, and sodic plagioclase with more or less muscovite and minor amounts of biotite, black tourmaline, or

garnet. Many pegmatites carry considerable albite (variety cleveländite) and small amounts of beryl, lithium tourmaline, lithium mica, and spodumene as well as any of over a hundred rarer species. Beautifully formed crystals of gem quality may line cavities or pockets in the pegmatite mass.

Texture and structure. One of the most characteristic features of pegmatite is the coarseness of grain. Feldspar and quartz crystals several feet long are not uncommon. Mica crystals 10 ft across, beryl crystals twice as long, and spodumene crystals nearly 50 ft long (Black Hills, S.Dak.) have been found. Equally characteristic is the great variation in grain size over short distances.

Much quartz and microcline perthite are intergrown in cuneiform fashion to produce graphic granite. *See* IGNEOUS ROCKS.

Zones. Zones of different minerals or textures give many pegmatites a banded appearance. Zones are roughly parallel to pegmatite margins and appear to have formed in succession from the walls inward. Outermost zones are fine grained and a few inches thick. Inner zones become thicker and coarser. Centrally located is an irregular lens-shaped core. Outer zones are more regular and continuous. Mineral associations within successive zones follow a definite sequence. Eleven associations have been recognized. No pegmatite, perhaps, exhibits all associations, but the sequence of associations is always followed (Fig. 1).

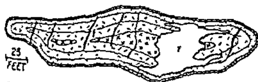


Fig. 1. Cross section through a zoned granite pegmatite idealized to show concentric zones, fracture fillings, and replacement body.

Fracture fillings. Fillings in fractures appear to cut across some pegmatites. These later bodies are composed mostly of quartz with small amounts of microcline, plagioclase, and mica. They range up to 100 ft long and 10 ft thick. Some run parallel to major pegmatite zones, but most cut across one or more zones (see Fig. 1).

Replacement bodies. These bodies are significant units in many pegmatites. They consist of quartz, albite (much as cleveländite), and muscovite in addition to numerous uncommon accessories. They appear to take the place of portions of the older pegmatite zones.

Replacement bodies are abundant, widespread, and commonly very large (up to several hundred feet long and tens of feet wide). They may form as tabular, lenticular, podlike, veinlike, or irregular bodies. Many replace along one or more zones of the main pegmatite. Others are clearly cross-cutting bodies (Fig. 1).

Origin. The origin of pegmatites is still unsolved. Many are believed to have formed from residual fluids of crystallizing rock melts (magmas). Crystallization of silicate minerals from a granitic magma enriches the residual fluid in silica, alkalies, volatiles (water, chlorine, fluorine, etc.), and many rare elements (boron, phosphorus, beryllium, lithium). From such fluids, material to build pegmatites may be derived. The high water content is believed to reduce greatly the viscosity of the fluid, permitting the growth of large crystals (coarse texture). Pegmatitic fluids forced out of crystallizing granitic magmas could be injected into fractures and openings in surrounding solid rock to crystallize and form zoned pegmatites.

Subsequently, fracture fillings and replacement bodies may form. Much of the replacing material may represent recrystallized earlier pegmatite minerals, but some may be metasomatic and derived from still later fluids.

Residual fluids not forced out of the crystallizing magma would be free to accumulate in pockets in the parent rock and solidify as segregation pegmatites with irregular and gradational boundaries (Fig. 2*a*). Gabbroic pegmatites commonly form along banded structure in gabbro (Fig. 2*b*). *See* GABBRO; MAGMA; METASOMATISM.

A commonly accepted theory supposes many pegmatites are of metamorphic origin, the materials being derived from the enclosing rocks and concentrated in zones of fracture and lower pressure (Fig. 2*c*). This mechanism, generally considered to explain various metamorphic veins, pods, and irregular bodies, involves transfer of material through solid rock, perhaps by the aid of intergranular fluids. *See* METAMORPHIC ROCKS; METAMORPHISM.

The formation of pegmatite on a grand scale, as seen in the Precambrian rocks of Canada and

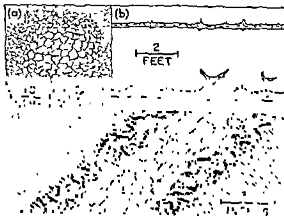


Fig. 2. Cross sections through pegmatites. (*a*) Segregation pegmatite in biotite granite. The core of coarse-textured quartz and microcline gradually passes outward to finer-grained granite. (*b*) Gabbroic pegmatite veins in gabbro. (*c*) Metamorphic pegmatite in biotite gneiss showing vague, irregular walls with biotite-rich selvedges in the gneiss. Central zone of pegmatite composed of quartz (clear) and potash feldspar. Marginal zones composed of quartz, plagioclase, and biotite (black).

Fennoscandia, may be due to regional metamorphism and granitization. See GRANITIZATION. [C.A.CA.]

Peking man

One of the best-known extinct human types, known from remains representing about 45 individuals. These were recovered from a breccia-filled cave fissure at Choukoutien, southeast of Peiping, China. The type was named *Sinanthropus pekinensis* by D. Black on the basis of a molar tooth found in 1927; the main finds were made from 1928 to 1937, largely under F. Weidenreich. All the material disappeared during an attempted evacuation to the United States at the outbreak of war in 1941. The cave contained a middle Pleistocene fauna indicating a Second Interglacial date, a stone chopper tool culture (the Choukoutienian), fire hearths, hackberry seeds, and bone tools used without prior shaping.



Reconstruction of a female *Sinanthropus* skull (somewhat modified after Weidenreich). *(After Weidenreich, from M. F. Ashley Montagu, An Introduction to Physical Anthropology, 2d ed., Charles C Thomas, 1951)*

The material indicates cannibalism rather than burial or accident, because it consists largely of teeth, jaw fragments, and skulls without faces or base portions. Rare parts of the postcranial skeleton reveal no essential differences from modern man. The brain case was thick, with a massive basal and occipital torus structure and heavy brow ridges. Average cranial capacity was 1075 cm^3, lower than other human types except Java man Peking man was evidently a somewhat advanced relative of the latter. See FOSSIL MAN. [W.W.H.]

Bibliography: F. Weidenreich, The skull of *Sinanthropus pekinensis*; a comparative study on a primitive hominid skull, *Paleontologia Sinica*, no. 127, 1943

Pelecaniformes

An order of aquatic birds characterized by having all 4 toes joined by webs. Six living and several fossil families are usually recognized, including such groups as the tropic birds, pelicans, gannets, frigate birds, cormorants, and anhingas or snake birds. All are primarily fish eaters, but methods of catching prey vary from the underwater pursuit of the long-necked snake bird (*Anhinga anhinga*) to the group fish drives of several pelicans (*Pelecanus*) and cormorants (*Phalacrocorax*). Although chiefly found in warmer waters, both fresh and salt, the Pelecaniformes have representatives among the gannets and cormorants in both far northern and far southern seas. Colonial nesting is highly developed in this order. See AVES. [K.C.P.]

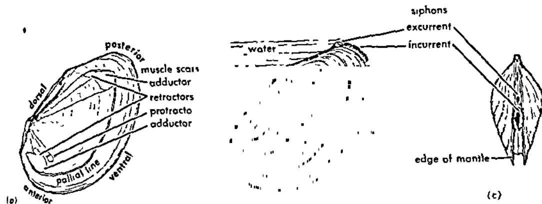
Pelecypoda

One of the large classes of the phylum Mollusca which contains the clams, oysters, and other bivalves. This class is world-wide in distribution, found in all oceans from the upper tide line to the greatest depths, as well as in most rivers, lakes, ponds, and other bodies of fresh water. The head has been reduced to a simple mouth without a radula or buccal mass, and is surrounded by labial palps. Feeding is accomplished by the gills in most groups; that is, the fine particles of food brought in through the incurrent siphon are captured by the mucus on the gills By means of cilia, these particles of food are led into food tracts and then carried to the mouth. The labial palps sort out the food elements and reject the sand and other unwanted material.

Morphology. In most primitive Protobranchia, such as *Nucula*, the gills do not collect the food; this function is performed by extensions of the labial palps which emerge from the shell and collect food material on the soft bottom The food is carried in by the cilia and sorted on the palps

In general, the pelecypods are bilaterally symmetrical animals. There are two calcareous valves to the shell which are produced by the right and left lobes of the mantle The valves are joined dorsally by a ligament, which acts as a spring, forcing the valves to gape The valves are pulled together by one or two transverse adductor muscles which are attached to the inner surfaces of the valves.

The visceral mass is mainly in the dorsal portion of the body and is attached to the dorsal and inner surface of the valves. Behind the mouth there is a short esophagus, a bulbous stomach, and a coiled intestine. The anus is posterior and opens into the mantle cavity near the excurrent siphon. The gills are suspended within the mantle cavity on each side of the foot. The foot is an extension of the visceral mass and is a muscular organ which can be extended beyond the ventral margin of the valves, used for burrowing or locomotion In many groups there is a gland which produces horny threads, forming the byssus which is used for attachment.



Anodonta, a pelecypod, fresh-water clam. (*a*) Inside of right valve (*b*) Left side (*c*) Posterior (*T I. Storer and R L Usinger, General Zoology, 3d ed , McGraw-Hill, 1957*)

The nervous system consists of three pairs of ganglia which have connecting fibers The heart is dorsal and in most pelecypods the intestine passes through it. The vascular system is incompletely closed.

Shell. The shell is composed of three layers. The outer layer, the periostracum, consists of a chitinlike material which protects the calcareous portion of the shell from acid action The middle layer is called the prismatic layer and is composed of calcareous crystals formed vertically or at a right angle to the innermost or laminated layer Both the prismatic and laminated layers are produced within a framework of conchiolin, a substance identical with the periostracum The external areas on the bivalve shell are called the posterior slope, the anterior slope, and the central region or the disk The dorsal margin at its highest point near the ligament is referred to as the umbo

Economic importance. Economically, the pelecypods are very important. Clams and oysters are a source of food for man in most portions of the world In addition, they are a food source for many bottom feeding fish, such as the cod, flounder, and haddock, which in turn, are important food fish for man On the negative side, the damage done by several species in the Teredinidae and Pholadidae to wooden ships, wharves, and other marine installations, costs many hundreds of millions of dollars each year. Some ...

... members of the Pholadidae can bore into wood, soft rock, and even poor grades of cement.

Most bivalves have free-swimming larvae which can be dispersed over wide areas by ocean currents Many species cause considerable damage as fouling organisms when they attach themselves to ships, buoys, and intake tunnels in electric plants. Cleaning off these mollusks is an exceedingly costly process

A few pelecypods produce nacreous shells which are called mother-of pearl The most noted are in the family Pteriidae, the pearl oysters Certain parasites invade the mantle of these mollusks and are eventually covered by a nacreous secretion. These are later pushed out of the thin tissue of the mantle and remain within the free area of the mantle cavity. In this group of bivalves, the hinge area rests on the substrate, whereas in most other bivalves the hinge area is uppermost As a consequence, the pearls remain within the mantle cavity Artificial pearl culture has become an important industry, particularly in Japan. *See* PEARL.

All of the fresh-water mussels in the family Unionidae produce nacreous shells and many species are used to make pearl buttons. Occasionally pearls of considerable value are produced by these fresh-water mussels In North America, fresh-water mussels once formed an important item of food for the American Indians *See* MOLLUSCA [W.J.C.]

Pelecypoda fossils

Pelecypods, popularly known as clams, are of ancient stock The earliest forms so far described were widely distributed by late Cambrian or early Ordovician time. Suitable environments for their development were available in the warm, shallow seas of Middle Cambrian time, but fossil pelecypods have not been found in rock strata of that age. Probably these mollusks diverged from the limpetlike Monoplacophorans during the Late Cambrian when calcification at two places in the shell formed two valves that were bridged by the chitinous outer layer or periostracum Differentiation was developed by Middle Ordovician time *See* MONOPLACOPHORA

Classification. No single basis for classification has been agreed upon by systematists. Paleontologists have favored various modifications of M. Neumayr's division—based upon the hinge—into five orders· Taxodonta, Dysodonta, Desmodonta, Palaeoconcha, and Heterodonta Zoologists prefer P Pelseneer's arrangement into four orders—based upon the gills (soft tissues not preservable in fossils) A synthesis of present-day inconsistent groupings may be expected in the *Treatise of In-*

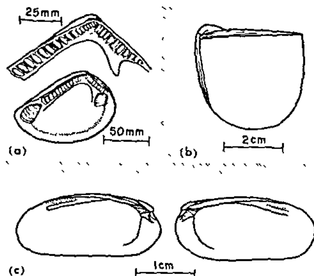


Fig 1. *(a)* Taxodont hinge, *Ctenodonta*, Middle Ordovician, Michigan *(after E. Ulrich, 1897)*. *(b)* Dysodont hinge, *Ambonychinia*, Upper Ordovician, Sweden *(after O. Isberg, 1934)*. *(c)* Heterodont hinge, *Permophorus*, Permian, Texas. Note presence of both cardinal and lateral teeth *(after N. Newell, 1957)*.

vertebrate Paleontology that will perhaps elaborate H. Douvillé's suggestion that dynamically there are three stocks, normal or free-moving forms, sedentary forms, and those adapted for burrowing.

Taxodonta. The earliest bivalves (Taxodonta) were smooth exteriorly, with a nacreous shell. The hinge (dorsal margin of contact between the valves) had a row of similar-shaped teeth, as in modern Nuculacea, and the chitinous bridge had become an elastic ligament (Fig. 1*a*).

Dysodonta. Next to appear, during the Ordovician, were the Dysodonta (ancestors of Mytilacea and Pteriacea), with a nearly toothless hinge and a ligament in grooves or pits (Fig. 1*b*).

Desmodonta and Palaeoconcha. The ill-defined groups Desmodonta (burrowers) and Palaeoconcha (simple, smooth-hinged forms) were relatively rare, and all pelecypods were outnumbered by Brachiopoda. Concentration of teeth along the center and ends of the hinge foreshadowed development of true heterodont dentition by Silurian time.

Heterodonta. Among some Heterodonta, the earlier nacreous shell structure gave place to crossed-lamellar or porcelaneous texture. Concentric ribbing developed sporadically, but radial ribbing was inconspicuous before Devonian time (Fig. 1*c*).

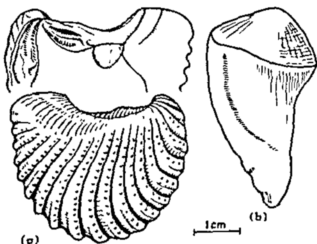


Fig. 2. *(a)* *Trigonia* (schizodont hinge), Upper Cretaceous, North Carolina *(after L. Stephenson, 1923)*. *(b)* A rudist, *Coralliochama*, Upper Cretaceous, California *(after C. White, 1885)*.

Geologic record. Not until the Carboniferous were pelecypods abundant enough to make good stratigraphic markers. From then on they largely displaced the Brachiopoda. Also during the Carboniferous, nonmarine pelecypods flourished, especially in the coal-forming swamps of Europe. *See* CARBONIFEROUS.

Although the Mesozoic pelecypods were eclipsed by their relatives, the Cephalopoda, their evolution continued. Most modern superfamilies can be traced to Jurassic or even to Triassic origins. Three distinctive Mesozoic groups are (1) *Inoceramus*, Jurassic to Cretaceous dysodonts with undulating concentric sculpture (Fig. 3*a*). Prisms of the nacreous shell are recognizable even when the shell is incomplete. (2) *Trigonia*, also Jurassic to Cretaceous, heterodont offshoots with divided, ser-

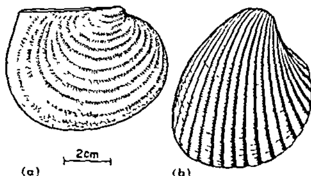


Fig. 3. *(a)* *Inoceramus*, Cretaceous, Texas *(after L. Stephenson, 1941)*. *(b)* *Venericardia*, Paleocene, California.

rate (schizodont) hinge teeth (Fig. 2*a*). (3) The rudists, mainly Cretaceous, aberrant heterodonts with one valve conical and up to 2 ft long, the other smaller, flattened to somewhat convex (Fig. 2*b*). Internal structures are complex

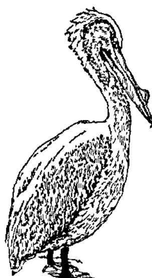
A few Tertiary genera originated as early as Cretaceous, but in most families, Late Cretaceous genera disappeared (sometimes suddenly), being replaced by new genera in the Paleocene. One Early Tertiary horizon marker is *Venericardia*, the classic "finger-post of the Eocene" (Fig. 3*b*).

Like all mollusks, pelecypods are sensitive to temperature and salinity differences. Thus, fossil forms are useful as indicators both of ecologic conditions and of time. Present-day emphasis is on assemblages rather than single guide species, for pelecypods tend to be wide-ranging. About 10,000 species have been described. [A.M.K.]

Bibliography: R. C. Moore (ed.), *Treatise of Invertebrate Paleontology*, Pt. N (in prep.); H. W. Shimer and R. R. Shrock, *Index Fossils of North America*, 1944

Pelican

Any of 10 species of the family Pelecanidae, a family of large aquatic, fish eating birds The pelican is characterized by a huge bill which is long, straight, and hooked at the tip, and opens into a large throat pouch There are two species in the United States The white pelican, *Pelecanus erythrorhynchos*, nests in North America from Great



The white pelican, *Pelecanus erythrorhynchos*, length to 5 ft (From E L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Slave Lake in Canada to California. It is white, with black tipped wings, and has a wingspread of as much as 9 ft The much smaller brown pelican, *P. occidentalis*, is entirely marine, nesting on both coasts of the United States See PELECANIFORMES [J D B]

Pellagra

A disease resulting from severe deficiency of niacin, a member of the vitamin B complex. Skin, gastrointestinal, and neurologic symptoms may occur. Wheat, milk, or egg proteins afford ample dietary requirements but corn protein may be deficient. This accounts for the regional, or endemic, form of pellagra in areas of restricted diet. See NIACIN

Redness, scaling, and brownish discoloration of the skin, particularly in areas exposed to sunlight, are common lesions. Marked enlargement of the tongue and drooling are prominent symptoms The mucous membranes of the mouth, eyes, urethra, and vagina may show swelling and have a bright red, smooth appearance. Gastrointestinal symptoms may be vague or may take the form of severe nausea, vomiting, and bloody diarrhea. Early nervous system involvement is displayed by the neurasthenic syndrome characterized by restlessness, anxiety, and insomnia. Organic psychoses may follow continued deficiency and are marked by memory loss, confusion, and a variable pattern of affective behavior, such as depression or paranoia. Actual delirium, limb rigidity, and certain uncontrolled reflexes mark the severe case See ABNORMAL BEHAVIOR.

Diagnosis may be difficult because of incompletely developed symptoms. In addition, pellagra is almost always accompanied by other vitamin B deficiencies of some degree. In recent years education and preventive medicine have largely eliminated pellagra from certain areas in the United States where it previously was prevalent. See VITAMIN. [E G ST.]

Pelmatozoa

One of the two subphyla of Echinodermata, comprising forms which typically are anchored to the substrate, at least during some part of the life history. Nearly all are extinct. The ovoid body is enclosed by polygonal calcite plates which collectively form the theca The mouth and anus both lie on the upper (adoral) surface, the gut thus assumes a U shape. The lower end of the theca (aboral pole) may develop a jointed stem, which is sometimes discarded by the adult The oldest known members date from the Lower Cambrian and had at that time diverged into two stocks, (1) the bilaterally symmetrical, stalked carpoids, and (2) the disk-shaped radially symmetrical edrioasteroids which had five food-collecting channels (ambulacral grooves) leading to the central mouth Cystids appeared in the Ordovician and soon adopted three-part (later five-part) radial symmetry; their ambulacral grooves became combined in various ways with flexible fingerlike organs called *brachioles*. In blastoids these features became stabilized into a rigid theca of unvarying pattern, whereas crinoids showed freedom of variation and acquired a crown of arms to carry the ambulacral grooves Crinoids exist today but no other Pelmatozoa are known to have survived the Paleozoic. See BLASTOIDEA; CARPOIDEA, CRINOIDEA; CYSTIDEA, ECHINODERMATA; EDRIOASTEROIDEA; EOCRINOIDEA; PARACRINOIDEA [H.B F.]

Bibliography: R. C. Moore, Pelmatozoa, *Bull Museum Comp. Zool Harvard*, vol. 112, 1954

Pelton wheel

An impulse type of hydraulic turbine. In the impulse turbine, pressure of the water supply is con-



A 12-in Pelton wheel (Photograph by R. L. Daugherty and P Kyropoulos)

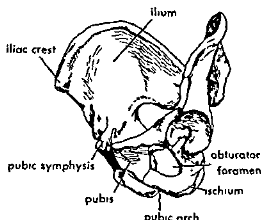
verted into velocity by a nozzle. The water jet then impinges on the buckets of the turbine wheel or runner. In a Pelton wheel the buckets have a splitter in the middle to divide the water jet, and cause it to flow across the cupped faces of the buckets and emerge at the sides. The water jet thus imparts its kinetic energy to the buckets, as illustrated Pelton wheels are usually operated from high-head sources. *See* HYDRAULIC TURBINE.

[F.H.R.]

Bibliography: R. L. Daugherty and A. C. Ingersoll, *Fluid Mechanics*, 5th ed., 1954.

Pelvis

The bony basin formed by the paired hip bones and the posterior sacrum and coccyx of the vertebral column. In childhood, each hip bone is in three separate parts, the lateral flat ilium, the anterior pubis, and the inferior ischium. These fuse and the hip socket becomes firm bone. The hip bones meet in front at a fixed joint, the pubic symphysis. Pel*vic viscera include bladder, rectum, and internal* reproductive organs. The pelvic floor consists of



Pelvic girdle, front view. (W T. Foster, Anatomy, Foster Art Service)

urogenital and perineal structures, muscular layers surrounding and supporting its intestinal and urinogenital outlets The female pelvis is wider and more shallow and broad than that of the male. Various pelvic muscle groups act on back, abdomen, legs, and viscera

Similar structures are common to other mammals. In birds and lower vertebrates the pelvis may be less developed; it does not appear in fishes and certain reptiles.

[E.C.ST.]

Pelycosauria

A primitive order of mammal-like reptiles (subclass Synapsida). They are characterized by a temporal fossa that lies low on the side of the skull. The group is known from rocks of the upper Carboniferous and lower and middle Permian. Three suborders are included: Ophiacodonta, primitive, partially aquatic carnivores; Edaphosauria, lowland, terrestrial herbivores; and Sphenacodontia, advanced, active carnivores. Size range is from about 1 ft to over 20 ft in total length. The majority inhabited lowland, deltaic environments They are best known from Permian deposits in northern Texas. Late in the early Permian the sphenacodonts gave rise to more advanced mammal-like reptiles, the therapsids. *See* SYNAPSIDA; THERAPSIDA.

[E.C.O.]

Pendulum

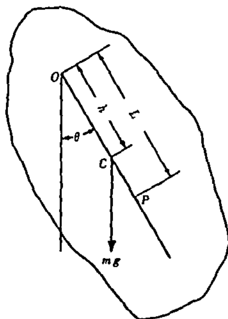
A rigid body mounted on a fixed horizontal axis, about which it is free to rotate under the influence of gravity. The period of the motion of a pendulum is virtually independent of its amplitude, and depends primarily on the geometry of the pendulum and on the local value of g, the acceleration of gravity. Pendulums have therefore been used as the control elements in clocks, or inversely as instruments to measure g.

Pendulum motion. In the schematic representation of a pendulum shown in the figure, O represents *the axis, and C the center of mass* The line OC makes an instantaneous angle θ with the vertical. In rotary motion of any rigid body about a fixed axis, the angular acceleration is equal to the torque about the axis divided by the moment of inertia I about the axis If m represents the mass of the pendulum, the force of gravity can be considered as the weight mg acting at the center of mass C. Therefore, the angular acceleration α is determined by the relation

$$-mgh \sin \theta = I\alpha = I\, d^2\theta/dt^2 \qquad (1)$$

where h is the distance OC, and t represents time.

If the amplitude of motion is small, $\sin \theta \approx \theta$ in radian measure. In this approximation, the motion is simple harmonic (*see* HARMONIC MOTION). The equation



Schematic diagram of a pendulum O represents the axis, C is the center of mass, P the center of oscillation.

$$-mgh\theta = I\, d^2\theta/dt^2 \tag{2}$$

has for its solution

$$\theta = A \sin(\omega t - \delta) \tag{3}$$

where the amplitude A and the phase δ are arbitrary constants. The angular frequency ω is given by

$$\omega^2 = mgh/I \tag{4}$$

The period T, time for a complete vibration (for [illegible] right to [illegible] nd fre- [illegible] me, are given by

$$T = 1/f = 2\pi/\omega = 2\pi\sqrt{I/mgh} \tag{5}$$

The actual form of a pendulum often consists of a long, light bar or a cord that serves as a support for a small, massive bob The idealization of this form into a point mass on the end of a weightless rod of length L is known as a simple pendulum. An actual pendulum is sometimes called a physical or compound pendulum In a simple pendulum, the lengths h and L become identical, and the moment of inertia I equals mL^2. Equation (5) for the period becomes

$$T = 2\pi\sqrt{L/g} \tag{6}$$

Because the value of g in metric units (about 9 8 m/sec²) is very nearly equal to π^2, a simple pendulum 1 meter in length has a period very close to 2 sec; the time for a single swing from right to left or left to right is approximately 1 sec.

Center of oscillation. Equation (6) can be used to define the equivalent length of a physical pendulum. Comparison with Eq (5) shows that

$$L = I/mh \tag{7}$$

The point P on line OC of the figure, whose distance from the axis O equals L, is called the center of oscillation. Points O and P are reciprocally related to each other in the sense that if the pendulum were suspended at P, O would be the center of oscillation

The proof of this relation follows from the parallel axis property of moments of inertia If the moment of inertia of the pendulum about its center of mass is equal to

$$I_0 = mb^2$$

then

$$I = m(h^2 + b^2)$$

and, by Eq (7)

$$hL = h^2 + b^2 \tag{8}$$

For a given value of b (the radius of gyration about the center of mass) and L, h is determined by Eq (8) to be either of the quantities

$$h = L/2 \pm \sqrt{(L^2/4) - b^2}$$

The sum of these two values equals L, so if one value of h is the distance OC, then CP must represent the other value of h that will give the same equivalent length.

If some particular body with a definite value for b is to be mounted about an arbitrary axis to make a pendulum, Eq (8) shows that L can never be less than $2b$, and that L will have this minimum value (and the period T will be a maximum) if h is made equal to b.

Center of percussion. The points O and P share another reciprocal property. If the body is free to move in the plane of the figure, instead of fixed on an axis, and an impulsive force is applied to the body at O, the initial motion of the body will be a rotation about P. For this reason, P is sometimes called the center of percussion about O.

If the motion of a pendulum is not limited to small amplitudes, Eq (2) is not an adequate substitute for the correct Eq (1). The angular velocity, $d\theta/dt$, can be derived as a function of displacement by multiplying both sides of Eq. (1) by $d\theta/dt$, and then integrating. The result, which can also be obtained directly from the principle of conservation of energy, is that

$$(d\theta/dt)^2 = 2\omega^2(\cos\theta - \cos\theta_0) \tag{9}$$

where θ_0 is the maximum displacement, or amplitude of the motion. Here ω is still the characteristic constant of the pendulum defined by

$$\omega^2 = mgh/I = g/L$$

although the relation between ω and frequency is no longer as simple as in the approximate Eq. (5).

To obtain θ as a function of time, introduce an angle ψ by the relation

$$\sin\psi = (1/k)\sin(\theta/2) \tag{10}$$

where

$$k = \sin(\theta_0/2) \tag{11}$$

Equation (9) becomes

$$\omega\, dt = \pm\frac{d\theta}{\sqrt{2(\cos\theta - \cos\theta_0)}} = \pm\frac{d\theta}{2\sqrt{\sin^2(\theta_0/2) - \sin^2(\theta/2)}} = \pm\frac{d\psi}{\sqrt{1 - k^2\sin^2\psi}}$$

If time is chosen zero when θ is zero,

$$\omega t = F(k,\psi) \tag{12}$$

where $F(k,\psi)$ is the standard elliptic integral of the first kind,

$$F(k,\psi) = \int_0^\psi \frac{dz}{\sqrt{1 - k^2\sin^2 z}}$$

Conversely, the angle θ can be expressed as elliptic function of time:

$$\sin(\theta/2) = k\,\mathrm{sn}\,(\omega t)$$

The accurate expression for the period T, obtai[ned] from Eq. (12), can be written in terms of the co[m]plete elliptic integral of the first kind, $K(k)$, a[s]

$$\omega T = 4F(k,\pi/2) = 4K(k) \qquad (13)$$

Numerical values for the ratio of the period for amplitude θ_0 to the period for infinitesimal amplitude are listed in the table.

Ratio of the period for amplitude θ_0 to the period for infinitesimal amplitude

θ_0	$T(\theta_0)/T(O)$	θ_0	$T(\theta_0)/T(O)$
0	1 0000	100°	1 2322
20°	1 0077	120°	1.3729
40°	1 0313	140°	1.5944
60°	1 0732	160°	2.0075
80°	1 1375	180°	∞

Pendulum types. The most important types of gravity pendulums are described in the following paragraphs.

Kater's reversible pendulum. This type is designed to measure g, the acceleration of gravity. It consists of a body with two knife-edge supports on opposite sides of the center of mass as at O and P (and with at least one adjustable knife-edge). If the pendulum has the same period when suspended from either knife-edge, then each is located at the *center of oscillation of the other, and the distance* between them must be L, the length of the equivalent simple pendulum. The value for g follows from Eq. (6) or Eq (13).

Ballistic pendulum This is a device to measure the momentum of a bullet. The pendulum bob is a block of wood, into which the bullet is fired The bullet is stopped within the block, and its mo*mentum transferred to the pendulum. This mo*mentum is determined from the amplitude of the pendulum swing *See* BALLISTICS, INTERIOR.

Spherical pendulum. This is a simple pendulum mounted on a pivot, so that its motion is not confined to a plane. The bob then moves over a spherical surface. A Foucault pendulum is a spherical pendulum suspended so that its plane of oscillation *is free to rotate Its purpose is to demonstrate the* rotation of the earth. If such a pendulum were mounted at the North Pole, the rotation of the earth under the pendulum would make it appear to a terrestrial observer that the plane of the pendulum's motion rotated 360° once every day. The plane of motion of a Foucault pendulum set up at a lower latitude rotates at a reduced rate, pro*portional to the sine of the latitude.*

Torsional pendulum. Despite its name, a torsional pendulum is not a pendulum. It is an example of a torsional harmonic oscillator, consisting of a disk or other body of large moment of inertia mounted on one end of a torsionally flexible rod. The other end of the rod is held fixed. If the disk is twisted and released, the torsional pendulum oscillates *harmonically.* Gravitation plays no part in its motion. For further details on the torsional pendulum, *see* HARMONIC MOTION. *See also* CLOCK; DIMENSIONAL ANALYSIS; SCHULER PENDULUM.

[J.M.KE.]

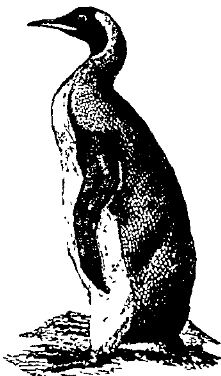
Bibliography: R. A. Becker, *Introduction to Theoretical Mechanics*, 1954.

Penetrance, gene

The percentage of the carriers of a gene which manifest its phenotypic effect. For example, when the gene for nicked wing in *Drosophila melanogaster* is homozygous, only 3% of the flies will show this character and the gene is said to have a penetrance of 3%. Nonmanifesting carriers of a gene are known as normal overlaps. The failure of a *gene to manifest itself may be due to specifiable* conditions of the environment, such as diet and temperature; to accidents of development such as developmental noise; or to interaction with other genes, that is, to the genetic background Only in the last case can the degree of penetrance be influenced by selection. In practice, genes with low penetrance are difficult to distinguish from the ef*fects of complex genetic situations. See* EXPRESSIVITY, GENE; GENE; GENETICS. [H.CR.]

Penguin

Any of about 20 species of marine, flightless birds of the order Sphenisciformes, found in the Southern Hemisphere. Except for one species on the Galapagos Islands, they are confined to the extreme *southern portion of the globe. Penguins are covered* by a thick layer of fat and are remarkably well adapted for life on the Antarctic ice cap and the adjacent barren areas. All their toes are directed forward and they walk upright when on land. Penguins catch fish, squid, and other animals by swimming underwater, using their flipperlike wings to propel themselves through the water rapidly. *They are all colonial The single egg of most spe-*



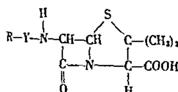
The penguin, Aptenodytes sp; length to 3 ft. (From P. M. Duncan, ed., Cassell's Natural History, Cassell, 1883)

ties is carried on top of the toes and incubated during the Antarctic winter. Penguins are of some value for oils and in the production of guano See SPHENISCIFORMES [J D B]

Penicillin

An antibiotic that is active against gram-positive bacteria and a few gram-negative ones. Penicillin, discovered in 1928 by Sir Alexander Fleming, was the first antibiotic to be widely and successfully used in the systemic therapy of acute bacterial infections in man. The phenomenal lack of toxicity to patients and the commanding therapeutic effectiveness of penicillin in a wide variety of infections are reflected in the dominant role it has held since 1940 in antibiotic therapy. In 1958 the production of penicillin in the United States for medical and veterinary use was 516,000 lb with sales of $60,321,000; 170,000 lb per year have been used in animal feeds See ANTIBIOTIC.

Chemistry. Penicillin is, strictly speaking, not any one substance but is the name given, collectively, to salts of a series of organic acids (pKa about 2.8) which may be considered to be *N*-acylated derivatives of 6-aminopenicillanic acid (see table). The latter contains the β-lactam-thiazolidine ring system common to all penicillins Many varieties of penicillin have been prepared by biosynthesis or by a combination of biosynthesis and chemical synthesis, and a few completely by chemical synthesis Most penicillin acids, being hydrophobic, can be extracted readily into a wide variety of organic solvents as the free acid and back into water as a salt.



Structure of penicillin

Most penicillins in aqueous solution are very unstable under acid (below pH 5.5) or alkaline (above pH 8.0) conditions, especially at raised temperatures (over 37°C). Phenoxymethyl penicillin, phenoxyethyl penicillin, and certain others are relatively stable in acid solution (half-life at pH 3 of many hours). Dry crystalline salts of all forms are fairly stable (100°C for many hours).

Commonly used salts of penicillin are K, Na, and procaine Most salts are readily soluble in water, low molecular weight alcohols and in aqueous mixtures of certain higher molecular weight alcohols; they are insoluble in ether, acetone, chloroform, and benzene. Salts with low water solubility, such as procaine salt, can be used as a slowly mobilized reservoir when injected intramuscularly at 3- or 4-day intervals.

Assay. Penicillin is readily assayed microbiologically by plate diffusion assay using *Staphylococcus aureus* or in broth by a turbidimetric assay. Higher concentrations can be chemically assayed. Concentrations of penicillin are expressed as in-

Structural relationship of the penicillins

Letter designation	Name	Y	R	Activity sodium salt, international units/mg
	6-Aminopenicillanic acid	H—	Absent	2
G	Benzylpenicillin	$-\overset{|}{C}=O$	$C_6H_5-CH_2-$	1667
X	p-Hydroxybenzylpenicillin	$-\overset{|}{C}=O$	$HO-C_6H_4-CH_2-$	900
F	2-Pentenylpenicillin	$-\overset{|}{C}=O$	$CH_3CH_2CH=CHCH_2-$	1600
K	n-Heptylpenicillin	$-\overset{|}{C}=O$	$CH_3(CH_2)_4CH_2-$	2300
N	D-4-Amino-4-carboxy-n-butylpenicillin (Synnematin B)	$-\overset{|}{C}=O$	$HC(NH_3^+)(COO^-)-CH_2-CH_2-CH_2-$	50
V	Phenoxymethylpenicillin	$-\overset{|}{C}=O$	$C_6H_5-O-CH_2-$	1670
	α-Phenoxyethylpenicillin	$-\overset{|}{C}=O$	$C_6H_5-O-CH(CH_3)-$	None
	6-Benzylsulfamidopenicillanic acid	$-S(=O)_2-$	$C_6H_5-CH_2-$	None

ternational units, each equivalent to 0 6 μg of pure benzylpenicillin. See BIOASSAY; STAPHYLOCOCCUS.

Antimicrobial activity. Penicillin is active (0 001 units/ml–5 units/ml for inhibition of growth) in general against the gram-positive bacteria, including (except for resistant strains) most species of such genera as *Streptococcus, Staphylococcus, Micrococcus, Clostridium, Borrelia, Corynebacterium* and *Bacillus*. The genera *Treponema, Neissera*, and *Actinomyces* are also sensitive. With a few notable exceptions (for example *Neisseria*), penicillin has no activity or low activity against gram-negative bacteria, including such genera as *Escherichia, Aerobacter, Klebsiella, Pasteurella, Eberthella, Pseudomonas, Vibrio, Brucella, Hemophilus, Mycobacterium*; yeasts, molds, and viruses are also resistant.

Serious and widespread epidemics among surgical patients and infants in hospitals have been caused by penicillin-resistant penicillinase-producing staphylococci. The enzyme penicillinase inactivates penicillin. These strains are thought to be naturally resistant since strains with induced in vitro resistance do not produce penicillinase. See BACTERIA; VIRUS; YEAST.

Pharmacology. The outstanding pharmacological characteristic of penicillin is its virtual nontoxicity (except for occasional allergic reactions, which can be fatal) to man and most animals.

Penicillin rapidly diffuses into blood after intramuscular or subcutaneous injection and is carried to nearly all tissues. Highly buffered forms and acid stable forms (such as phenoxymethyl penicillin) may be used for oral dosage to reduce the destruction by the acid in the stomach. The drug is rapidly excreted by the kidneys and, therefore, the dose must be renewed every 3–4 hours in order to maintain therapeutic blood levels unless some form of repository dosage, such as the procaine salt (dosage once in 3 or 4 days), is used

Therapeutic use. Dosages range from a low total of 50,000 units for gonorrhea to 1–3 million units per day for 20–30 days or more for bacterial endocarditis Some of the other diseases which may be successfully treated with penicillin are pneumococcal pneumonia, empyema, cellulitis, gas gangrene, meningicoccal meningitis, syphilis, dental and oral infections, osteomyelitis, a wide variety of infected wounds (both accidental and surgical) when these infections are caused by penicillin-sensitive bacteria. Penicillin is not effective for therapy of infections caused by viruses, by resistant gram-negative bacteria, or by resistant strains of organisms generally penicillin-sensitive. See GANGRENE, GAS; GONORRHEA; MENINGITIS; PNEUMONIA; SYPHILIS.

Biosynthesis. Penicillin occurs in a number of species of *Penicillium* and *Aspergillus*. The "natu- [illegible] lactose me- [illegible] with smaller [illegible] 5-aminopeni- [illegible] nulate when the medium is supplemented with phenylacetic acid, but it is acylated by the mold to benzylpenicillin, as are the "natural" penicillins under these conditions. The mold derives the benzyl radical from the phenylacetic acid and incorporates it intact into the penicillin molecule. A wide variety of penicillins can be produced when appropriate precursors are added to the medium, for example phenoxyacetate gives penicillin V, *N*-*p*-chloroacetyl-*dl*-valine gives *p*-chlorobenzylpenicillin.

6-Aminopenicillanic acid can be chemically acylated to give desirable new penicillins not capable of being produced by the organism (for example α-phenoxyethyl penicillin). Penicillin can also be chemically or microbiologically degraded and then built up to form new penicillins.

In commercial production, high-producing strains (induced mutants) of *Penicillium chrysogenum* are used. The inoculum for tank fermentation is started by using spores from 7-day mold growth on cracked corn to start a shake flask of cornsteep-lactose medium; this flask is used directly to start the inoculum tank stage. The inoculum tank may be used to inoculate the final fermentation tank of 5,000–20,000 gal

The medium for the final fermentation is composed of a nitrogen source, usually cornsteep liquor, a carbohydrate, formerly lactose but now usually glucose or sucrose, added slowly throughout the fermentation at a rate which will keep the pH from going much above 7 or below 6 5, a precursor, and an antifoam. The precursor is phenylacetic acid, or phenylacetamide if benzylpenicillin is the desired product; phenoxyacetic acid if phenoxymethyl penicillin is the desired product. The antifoams used are fats, such as corn oil and lard oil, or synthetic antifoams such as the silicones.

After inoculation of the final fermentation, it is aerated vigorously and mechanically agitated throughout the fermentation period of 100–140 hours. The temperature is carefully controlled to 24–25°C. The fermentation goes gradually through several stages: a growth stage, a penicillin production stage, and a final stage of reduced production.

Maximum titers of penicillin may reach 5,000 units/ml or more (titers of 10,000 units/ml are commonly rumored but such data are closely guarded industrial secrets). The first stage of recovery is filtration of the mycelium, followed by acidification to pH 1.5 and immediate extraction of the free acid into butyl or amyl acetate. Back extraction into an aqueous solution buffered with potassium salts may be followed by recrystallization from *n*-butanol as potassium benzylpenicillin, which may later be converted into the procaine salt. Mycelial residues containing some penicillin may be dried for use as animal feed materials. Phenoxymethyl penicillin is fermented and recovered by a method nearly identical to that just described.

[R.E.B.]

Bibliography: H T Clarke (ed), *The Chemistry of Penicillin*, 1949; Federal Trade Commission, *Economic Report on Antibiotics Manufacture*,

1938, A Fleming, *Penicillin, Its Practical Application*, 1946; H W. Florey et al, *Antibiotics*, 1949; J W Foster, *Chemical Activities of Fungi*, 1949; H S Goldberg, (ed), *Antibiotics, Their Chemistry and Non Medical Uses*, 1959; L. A. Underkofler and R J Hickey, *Industrial Fermentations*, 2 vols., 1954

Penis

The male organ of copulation, the phallus, which consists basically of three elongated masses of erectile tissue in the human The central corpus spongiosum lies below and in the groove formed by the paired corpora cavernosa The urethra runs along the underside of the spongiosum and then

forms the retractable foreskin, or prepuce. The organ is held firmly in place by fibrous tissue and ligaments that bind it to the under side of the pubic arch Erection is by nerve stimulation that causes engorgement of the spiral helicine arteries and the plentiful venous sinuses of the organ

Paired penes first appear in reptiles as modifications of the cloacal wall; turtles and crocodiles have a slightly erectile single organ In most birds no true penis is present, although internal fertilization is common through cloacal deposition Penes in mammals show much variation, but the essential features are those of the human penis. In some mammals, such as the dog, the organ is made more rigid by the presence of a penile bone, the os priapi. *See* COPULATORY ORGAN [E G ST.]

Pennatulacea

An order of the subclass Alcyonaria, commonly called the sea pens These animals lack stolons and live with their bases embedded in the soft substratum of the sea. The colony consists of a distal rachis bearing many polyps and a polypless proximal peduncle, whose terminal end sometimes expands to form a bladder The colony of *Pennatula* looks like a feather (Fig 1*a*), being formed of numerous secondary polyps which arise from leaf-shaped lateral expansions of the very elongated primary axial or terminal polyp In the other form of colony, the polyps arise directly from the primary one as in *Cavernularia* (Fig. 2), *Veretillum*, and *Renilla* (Fig 1*b*,*c*). The colony has a horny unbranched axial skeleton comprised of pennatulin and some calcium carbonate and phosphate, but *Cavernularia*, including the luminous species, has a rudimentary one, and in *Renilla* it is lacking Smooth calcareous spicules which are oval, rod-, spindle-, and needle-shaped are found in the coenenchyme.

The gastrovascular cavity of the primary polyp is divided into four principal canals which run longitudinally around the axial skeleton through a spongy tissue containing solenial networks (Fig. 3). The cavity of each secondary polyp is inter-

Fig 1. (a) Young colony of *Pennatula phosphorea* (after H. Jurgensen). (b) *Renilla amyethystina* Verrill (after W. Kükenthal). (c) *Veretillum cynomorium*.



Fig 2. *Cavernularia haberen* (preserved specimen)

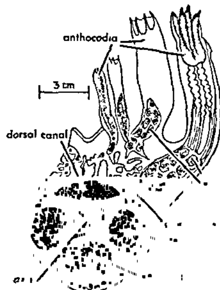


Fig 3. Cross section of *Funiculina quadrangularis* (after A. Kolliker).

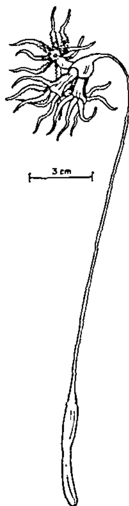


Fig. 4. *Umbellula encrinus* (after D. Danielssen)

connected by such canal systems. The musculatures are strongly developed, especially in the peduncle. Longitudinal muscles contract the body, and circular ones serve to force the water out of the main canals. In some littoral species, the polyp contracts or expands according to the tidal or diurnal and nocturnal rhythms.

All Pennatulacea are dimorphic. The two kinds of polyps resemble those of the other orders. In addition to the genera indicated above the order also includes *Virgularia*, *Umbellula* (Fig. 4), *Pteroeides*, and others. *See* ALCYONARIA.

[K.A.]

Pennsylvanian

An important division of late Paleozoic time, varyingly considered to rank as an independent period or as a main subdivision (termed epoch) of the

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

Carboniferous Period. It is named Pennsylvanian for outcrops of coal-bearing rock formations in Pennsylvania. These are steeply folded in the anthracite district in the northeastern part of the state but nearly flat-lying around Pittsburgh in the west.

Most American geologists class the Pennsylvanian rocks as a geologic system, whereas equivalent rocks out [illegible] designated as [illegible] ures. On the [illegible] is marked by such great coal making as distinguishes Pennsylvanian time, but marine deposits occur also *See* CARBONIFEROUS.

[R C.M.]

Pentaerythritol

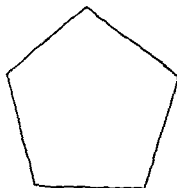
A white crystalline compound, $C(CH_2OH)_4$, which has four methylol groups arranged symmetrically around a central carbon atom. It is only slightly soluble in alcohols and other organic solvents but is moderately soluble in cold water and freely soluble in hot water. Pentaerythritol melts at 261–262°C.

Pentaerythritol, also called tetramethylolmethane, is made by reaction of formaldehyde and acetaldehyde in the presence of either calcium hydroxide or sodium hydroxide During World War II, appreciable quantities were used to make the explosive, pentaerythritol tetranitrate (PETN), now, most of it enters into the manufacture of alkyd resins and other coating compositions. In 1956, production of pentaerythritol in the United States was 57,000,000 lb. *See* ALCOHOL, POLYHYDROXY ALCOHOL.

[D W.K]

Pentagon

A geometric figure formed by the five line segments, or sides, that join in order five ordered points, or vertices, of a plane. In elementary geometry it is supposed that the sides do not cross, and even that the figure bounds a convex region of the plane (convex pentagon). The star-shaped pentagon (pentagram) was used as a symbol of recognition by the Pythagorean brotherhood. The construction



Regular pentagon.

there may be an electronic reaction. With moderately large signal voltages the third grid may become negative enough each cycle to repel low-velocity electrons approaching it from the oscillator section of the tube. These electrons are repelled into the oscillator section and constitute an [illegible] local oscill-

[illegible] between signal and oscillator circuits may be reduced by using a tube with a special electrode structure, which prevents electrons that are reflected back from entering the interaction space of the oscillator section. This is done by means of collector plates, which further increase the electrostatic shielding between the signal and oscillator circuits. The resulting operating characteristics are appreciably superior to those of the ordinary tube. *See* VACUUM TUBE. [E.H.K.]

Pentastomida

A class of blood-sucking, internal parasites belonging to the phylum Arthropoda. They are frequently referred [illegible] from 20 [illegible]



Adult female pentastomid (*Armillifer moniliformis*) in situ in lungs of reticulated python.



Head region of a pentastomid (*Porocephalus crotali*) from lungs of rattlesnake.

male. Fertilization is internal and the ripe eggs from the female ovary reach the outside by way of the mouth of the host. The intermediate host becomes infected by taking in eggs adhering to food or by drinking contaminated water. Young larvae, released from the eggs during digestion, encyst in a visceral organ. Later, as nymphs, they may reach a final host which has fed upon the infected intermediate host. Here, in the final host, the nymphs migrate to the lungs, become attached by their hooks, and grow to maturity.

Human infection frequently occurs in West Africa and Northern Europe, where man acts as an accidental intermediate host to the nymphal form. The liver is the most common site of infection Serious pathological conditions sometimes arise from the presence of these parasites within the human body.

The following points of similarity are advanced as evidence to relate the pentastomids with the arthropods: (1) Jointed appendages are present in the embryo; (2) the skin is provided with stigmata or breathing pores; (3) the reproductive system is highly developed, especially in the male; (4) ecdysis or molting of the skin occurs at frequent intervals in larvae and nymphs. *See* ARTHROPODA; CEPHALOBAENIDA; POROCEPHALIDA. [H.R.H.]

Bibliography: H. R. Hill, Annotated bibliography of the Linguatulidae, *Bull. So. Calif. Acad. Sci.*, 47, 1948; L. W. Sambon, A synopsis of the family Linguatulidae, *J. Trop Med. Hyg.*, 25, 1922; C. W. Stiles and A. Hassall, Key-catalogue of the Crustacea and Arachnoids of importance in public health, *Hyg. Lab Bull.* 148, 1927.

Pentlandite

A mineral having composition $(Fe,Ni)_9S_8$. Pentlandite is the major ore of nickel. It crystallizes in the isometric system, but crystals are rare. It is usually massive, showing a well-defined octahedral parting The hardness is 3.5–4 (Mohs scale) and the specific gravity varies from 4.6 to 5.0, depending on the ratio of iron to nickel; greater amounts of iron cause an increase in the specific gravity. The luster is metallic and the color yellowish bronze. Pentlandite is usually associated with pyrrhotite which it closely resembles in appearance but the two can be distinguished by the octahedral parting and lack of magnetism of pentlandite It is found at many localities in small amounts but its chief occurrence is at Sudbury, Ontario, where it is mined on a large scale as a nickel ore *See* NICKEL. [C.S.HU.]

Pentode, vacuum

A 5-electrode vacuum tube. The pentode is the most versatile and extensively used of all vacuum tubes. It is used for virtually all purposes for which vacuum tubes can be used. These functions include amplification, oscillation, mixing, pulse generation, and various timing, counting, and control circuits.

Electrode arrangement. The five electrodes in order are the cathode, control grid, screen grid, suppressor grid, and plate (*see* VACUUM TUBE). The pentode was developed from the tetrode by adding the suppressor grid to eliminate the exchange of secondary electrons between the screen grid and the plate. The suppressor grid is a coarse mesh grid through which the electrons from the cathode can readily pass to reach the plate. The suppressor grid is operated at cathode potential, causing a deep dip in the potential profile between

screen grid and plate and therefore inhibiting low-velocity secondary electrons emitted by the plate from reaching the screen grid. A typical set of potential profiles in a pentode is shown in Fig 1. Both the plate and the screen grid present negative gradients of potential to secondary electrons created at their surface. This eliminates the exchange of secondary electrons between screen and plate and results in current-voltage characteristics which are almost exactly those that would occur in a perfect screen-grid tube having no secondary emission. Further reference to Fig. 1 shows that the cathode, control grid, and screen grid have the same relative location and potential as in a triode. These three electrodes serve to produce a stream of electrons which passes on to the plate. The mutual conductance of pentodes is about the same as in triodes, but the amplification factor and dynamic plate resistance are much higher because of the shielding effect of two extra grids between the control grid and the plate.

Electrostatic field. A field plot of lines of equal potential within a plane-electrode pentode is shown in Fig 2. Some representative electron paths and a portion of a potential profile are also shown. This plot is for a zero control-grid potential, a high screen grid potential, a zero (usual) suppressor-grid potential, and a low plate potential The grid wires in a pentode are commonly not aligned, with the result that there is considerable dispersion of the electrons passing through the tube. In particular, electrons may be deflected by the suppressor grid wires to such an extent that they cannot reach the plate It is this effect which causes the plate current to go to zero as the plate voltage goes to zero. Secondary electrons are not shown in this figure because they are mostly low-velocity electrons which will be attracted back to the screen or plate where they were created. The potential profiles A and B are taken through the tube as shown,

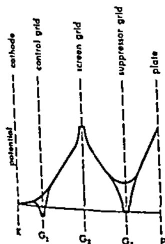
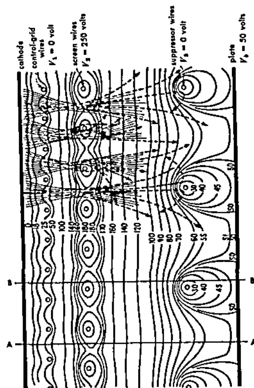


Fig 1 Variation of voltage between electrodes in a pentode (From K R Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)



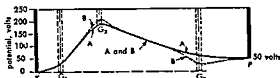


Fig. 2. Electrostatic field and electron paths in a pentode (*After Schade from K R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948*)

with A being midway between screen and suppressor wires, and B being closer to each of these.

Pentode characteristics. The current-voltage characteristics of a pentode are uniform and simple. The plate-current–plate-voltage characteristics are shown in Fig. 3 by the dashed curves. When the plate voltage is more than about half the screen voltage, the curves are practically constant, although they have a small positive slope. For low values of plate voltage, the plate current drops to zero. At zero grid voltage, the plate current is relatively high. As the control grid is made more negative, the plate current decreases

The solid curves in Fig. 3 are curves of total space current, which includes both plate current and screen current. These have the same general shape as the plate-current curves, although the magnitude is naturally larger The differences between these two sets of curves are the screen-current curves shown in Fig 4 These curves show that the screen current decreases as the plate voltage increases The decrease is very rapid at low plate voltages, because at zero plate voltage all of the current from the cathode goes to the screen. As the plate voltage is increased, the plate will collect



Adult female pentastomid (*Armillifer moniliformis*) in situ in lungs of reticulated python



Head region of a pentastomid (*Porocephalus crotali*) from lungs of rattlesnake.

male. Fertilization is internal and the ripe eggs from the female ovary reach the outside by way of the mouth of the host. The intermediate host becomes infected by taking in eggs adhering to food or by drinking contaminated water. Young larvae, released from the eggs during digestion, encyst in a visceral organ. Later, as nymphs, they may reach a final host which has fed upon the infected intermediate host. Here, in the final host, the nymphs migrate to the lungs, become attached by their hooks, and grow to maturity.

Human infection frequently occurs in West Africa and Northern Europe, where man acts as an accidental intermediate host to the nymphal form. The liver is the most common site of infection. Serious pathological conditions sometimes arise from the presence of these parasites within the human body.

The following points of similarity are advanced as evidence to relate the pentastomids with the arthropods: (1) Jointed appendages are present in the embryo; (2) the skin is provided with stigmata or breathing pores; (3) the reproductive system is highly developed, especially in the male; (4) ecdysis or molting of the skin occurs at frequent intervals in larvae and nymphs. *See* ARTHROPODA; CEPHALOBAENIDA; POROCEPHALIDA. [H.R.H.]

Bibliography: H. R. Hill, Annotated bibliography of the Linguatulidae, *Bull. So. Calif. Acad. Sci.*, 47, 1948; L. W. Sambon, A synopsis of the family Linguatulidae, *J. Trop. Med. Hyg.*, 25, 1922; C. W. Stiles and A. Hassall, Key-catalogue of the Crustacea and Arachnoids of importance in public health, *Hyg Lab. Bull.* 148, 1927.

Pentlandite

A mineral having composition $(Fe,Ni)_9S_8$. Pentlandite is the major ore of nickel. It crystallizes in the isometric system, but crystals are rare. It is usually massive, showing a well-defined octahedral parting. The hardness is 3.5–4 (Mohs scale) and the specific gravity varies from 4.6 to 5.0, depending on the ratio of iron to nickel; greater amounts of iron cause an increase in the specific gravity. The luster is metallic and the color yellowish bronze. Pentlandite is usually associated with pyrrhotite which it closely resembles in appearance but the two can be distinguished by the octahedral parting and lack of magnetism of pentlandite. It is found at many localities in small amounts but its chief occurrence is at Sudbury, Ontario, where it is mined on a large scale as a nickel ore. *See* NICKEL. [C.S.HU.]

Pentode, vacuum

A 5-electrode vacuum tube. The pentode is the most versatile and extensively used of all vacuum tubes. It is used for virtually all purposes for which vacuum tubes can be used. These functions include amplification, oscillation, mixing, pulse generation, and various timing, counting, and control circuits.

Electrode arrangement. The five electrodes in order are the cathode, control grid, screen grid, suppressor grid, and plate (*see* VACUUM TUBE) The pentode was developed from the tetrode by adding the suppressor grid to eliminate the exchange of secondary electrons between the screen grid and the plate. The suppressor grid is a coarse mesh grid through which the electrons from the cathode can readily pass to reach the plate. The suppressor grid is operated at cathode potential, causing a deep dip in the potential profile between

screen grid and plate and therefore inhibiting low-velocity secondary electrons emitted by the plate from reaching the screen grid. A typical set of potential profiles in a pentode is shown in Fig. 1. Both the plate and the screen grid present negative gradients of potential to secondary electrons created at their surface. This eliminates the exchange of secondary electrons between screen and plate and results in current-voltage characteristics which are almost exactly those that would occur in a perfect screen-grid tube having no secondary emission. Further reference to Fig. 1 shows that the cathode, control grid, and screen grid have the same relative location and potential as in a triode. These three electrodes serve to produce a stream of electrons which passes on to the plate. The mutual conductance of pentodes is about the same as in triodes, but the amplification factor and dynamic plate resistance are much higher because of the shielding effect of [illegible]

in Fig. 2. Some representative electron paths and a portion of a potential profile are also shown. This plot is for a zero control-grid potential, a high screen-grid potential, a zero (usual) suppressor-grid potential, and a low plate potential. The grid wires in a pentode are commonly not aligned, with the result that there is considerable dispersion of the electrons passing through the tube. In particular, electrons may be deflected by the suppressor grid wires to such an extent that they cannot reach the plate. It is this effect which causes the plate current to go to zero as the plate voltage goes to zero. Secondary electrons are not shown in this figure because they are mostly low-velocity electrons which will be attracted back to the screen or plate where they were created. The potential profiles A and B are taken through the tube as shown,

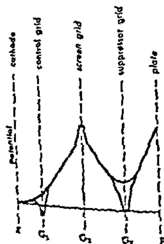


Fig. 1 Variation of voltage between electrodes in a [illegible] Spangenberg, Vacuum Tubes, [illegible]

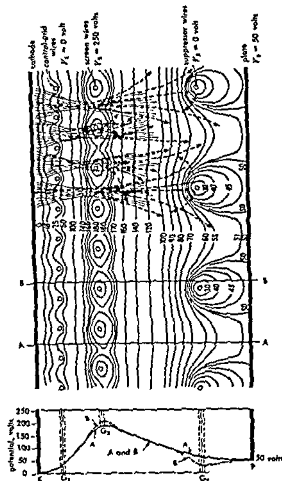


Fig. 2 Electrostatic field and electron paths in a pentode (After Schade from K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

with A being midway between screen and suppressor wires, and B being closer to each of these.

Pentode characteristics. The current-voltage characteristics of a pentode are uniform and simple. The plate-current–plate-voltage characteristics are shown in Fig. 3 by the dashed curves. When the plate voltage is more than about half the screen voltage, the curves are practically constant, although they have a small positive slope. For low values of plate voltage, the plate current drops to zero. At zero grid voltage, the plate current is relatively high. As the control grid is made more negative, the plate current decreases.

The solid curves in Fig. 3 are curves of total space current, which includes both plate current and screen current. These have the same general shape as the plate-current curves, although the magnitude is naturally larger. The differences between these two sets of curves are the screen-current curves shown in Fig. 4. These curves show that the screen current decreases as the plate voltage increases. The decrease is very rapid at low plate voltages, because at zero plate voltage all of the current from the cathode goes to the screen. As the plate voltage is increased, the plate will collect

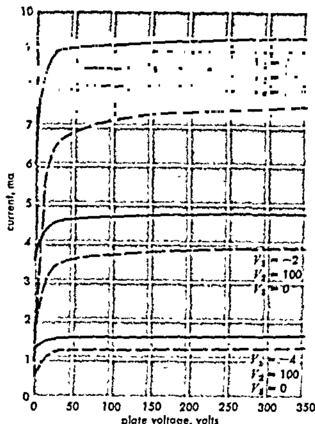


Fig. 3 Plate current-voltage characteristics of pentode (From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

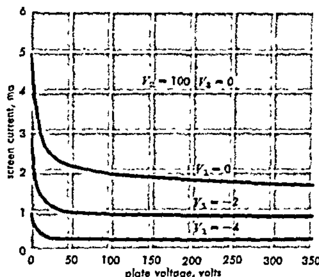


Fig 4. Screen current as a function of plate voltage in a pentode. (From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

the major portion of the cathode current with a certain fraction going to the screen because of direct interception of electrons. The screen current-plate voltage curves all have about the same shape, with the current decreasing as the grid voltage becomes more negative. A final representation of pentode characteristics is shown in Fig. 5. These are the transfer characteristics

Pentode tube coefficients. The same coefficients that are used for triodes are used for pentodes. These include the amplification factor, the mutual conductance, and the dynamic plate resistance. Of these three factors the two most commonly quoted are, in order of their interest, the mutual conductance and the dynamic plate resistance. The *amplification factor* of pentodes is seldom quoted, because it is so high that it has virtually no significance. The dynamic plate resistance is also high, commonly of the order of megohms, which results in the pentode's exhibiting the characteristics of a constant-current generator. The mutual conductance of a pentode will be similar in its range of values to that of triodes. The significance of the term constant-current generator is that the output current of the pentode will be essentially independent of the load resistance. This assumes what is generally true, namely that the load resistance is small compared to the dynamic plate resistance.

The mutual conductance of a pentode can be estimated by considering that the cathode, control grid, and screen grid constitute a triode to which the triode formulas can be applied. The mutual conductance obtained by this means needs only to be reduced by a fraction corresponding to the fraction of the total current intercepted by the screen grid. Typical values of mutual conductance are in the range of 5000–20,000 micromho. As is the case with triodes, the mutual conductance varies approximately as the cube root of the cathode current.

The dynamic plate resistance may be estimated from the reciprocal of the slope of the plate-current–plate-voltage characteristics. It is commonly in the range of several hundred thousand ohms to several megohms. The dynamic plate resistance tends to vary inversely as the cube root of the cathode current. Because the dynamic plate resistance is so high, the amplification of a pentode amplifier stage can be estimated closely by simply taking the product of the mutual conductance and the load resistance.

The amplification factor is relatively constant and may be estimated by considering that the pentode is really a combination of three triodes

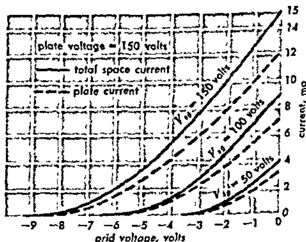


Fig. 5. Plate current and space current as a function of grid voltage in a pentode (From K. R Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

found by taking the three possible combinations of three adjacent electrodes The over-all amplification factor will be the product of the amplification factors of these three hypothetical triodes.

Plate-current cutoff. Pentodes may be classified according to their plate-current–grid-voltage characteristics

Sharp cutoff pentode. The so-called sharp-cutoff pentode is the more common form used for ordinary amplifier circuits The control grid of this tube has a uniform geometrical form over the entire length of the cathode. In a tube with such a grid, the plate-current–control-grid voltage characteristic, as shown in Fig 5, comes sharply to zero as the grid voltage is made more negative. This is the reason for the designation of this type of tube. The sharpness of the cutoff is of course a relative matter; that shown in Fig. 5 is relatively sharp as compared to the type to be described next.

Remote-cutoff pentode. Another type of pentode commonly used is the so-called remote-cutoff pentode This is also referred to as the variable-mu (μ) pentode In this type of pentode the control-grid structure is not uniform over the entire length of the cathode At one end the grid is constructed so that the amplification factor μ is relatively low, and the structure is then tapered to the other end where it is such that the amplification factor is relatively high. This can be done most simply by using a helical grid whose pitch decreases progressively from one end to the other, so that at one end the spacing between wires is large, giving rise to a low μ, whereas at the other end it is small, giving rise to a high μ With such a tube the current is cut off gradually as the control grid is made more negative. This happens because the high-μ end of the grid cuts off first, as the control-grid voltage is made more negative, and then gradually the entire grid cuts off more and more until the current is finally reduced to zero.

The structure described makes it possible to change the amplification of a pentode stage by changing the grid voltage. This characteristic is caused by the gradual change in mutual conductance with grid voltage over a large range of grid voltage. This tube is useful in automatic gain-control circuits. In these circuits a voltage is developed proportional to the signal strength and is fed back to the grids of some pentodes of the remote-cutoff or variable-mu type in such a direction as to keep the output constant. It is also a consequence of this type of operation that the distortion, or intermodulation, between two signals is reduced This results because the current characteristics of the pentode do not exhibit sharp curves or corners but have more gradual transitions, a characteristic that contributes to reduced distortion and intermodulation [K.R.S.]

Penumbra

That portion of a shadow illuminated by only part of a radiating source. A penumbra exists only when the radiating source, usually a light source, has an appreciable angular size. With a point source, the transition from light to dark at the edge of a shadow is abrupt. This is not so, however, with an extended source, where the opaque object forming the shadow may block off the radiation from only a part of the bright surface of the source. Depending on what fraction of this surface is exposed, the illumination in the penumbra varies from zero at the edge of the full shadow to the maximum where the entire source is exposed. See SHADOW; *see also* ECLIPSE, ASTRONOMICAL; UMBRA. [F.A.J.]

Pepper

The common garden pepper (*Capsicum annuum*), a warm-season perennial of American origin which belongs to the plant order Tubiflorales. This species includes all the peppers grown in the United States with the exception of the variety Tabasco (*C. frutescens*). Other species, C. pubescens and C. pendulum, are grown principally in South America. None are related to *Piper nigrum*, the woody plant from which black and white pepper are obtained

Sweet, or nonpungent peppers, generally picked when immature, are commonly cooked or eaten raw in salads. Popular varieties are California Wonder and Yolo Wonder. Perfection is the most popular pimento variety grown for canning

Hot or pungent peppers are most often harvested when ripe and ground into powder for seasoning, although some are canned The pungent compound, capsaicin, is concentrated mainly in the placental or seed tissue. Popular varieties are Anaheim, Cayenne, and Chili

Propagation is by seed with plants started in greenhouses or outdoor beds and transplanted to the field after 6–10 weeks Field spacing varies; plants 18–24 in. apart in 30–36-in. rows are common. Long warm seasons favor high yields and quality; however, high temperatures (above 90°F) and low humidity inhibit normal fruit set.

Harvesting of green sweet peppers begins when the fruit are near full-size but before they become mature and turn red or yellow, usually 60–80 days after field planting. Hot peppers are picked fully ripe for drying, generally 70–90 days after field planting Florida, Georgia, and New Jersey are important sweet-pepper producing states The total annual farm value in the United States is approximately $22,000,000. *See* PAPRIKA; PIMENTO; TUBIFLORALES, VEGETABLE GROWING. [H.J.C.]

Pepper (black)

One of the oldest and most important of the spices. It is the dried, unripe fruit of a weak climbing vine, *Piper nigrum*, a member of the pepper family (Piperaceae), and a native of India or Indomalaysia The fruits are small one-seeded berries which, in ripening, undergo a color change from green to red to yellow. When in the red stage, they are picked, sorted, and dried. The dry, wrinkled berries (peppercorns) are ground to make the familiar black pepper of commerce. White pepper is obtained by grinding the seed separately from

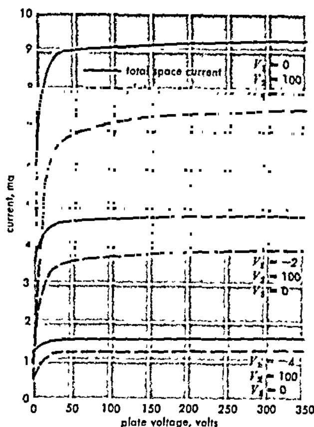


Fig. 3. Plate current-voltage characteristics of pentode. (From K R Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

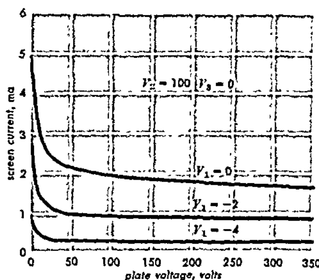


Fig 4 Screen current as a function of plate voltage in a pentode (From K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

the major portion of the cathode current with a certain fraction going to the screen because of direct interception of electrons. The screen current-plate voltage curves all have about the same shape, with the current decreasing as the grid voltage becomes more negative. A final representation of pentode characteristics is shown in Fig. 5. These are the transfer characteristics.

Pentode tube coefficients. The same coefficients that are used for triodes are used for pentodes. These include the amplification factor, the mutual conductance, and the dynamic plate resistance. Of these three factors the two most commonly quoted are, in order of their interest, the mutual conductance and the dynamic plate resistance. The amplification factor of pentodes is seldom quoted, because it is so high that it has virtually no significance. The dynamic plate resistance is also high, commonly of the order of megohms, which results in the pentode's exhibiting the characteristics of a constant-current generator. The mutual conductance of a pentode will be similar in its range of values to that of triodes. The significance of the term constant-current generator is that the output current of the pentode will be essentially independent of the load resistance. This assumes what is generally true, namely that the load resistance is small compared to the dynamic plate resistance.

The mutual conductance of a pentode can be estimated by considering that the cathode, control grid, and screen grid constitute a triode to which the triode formulas can be applied. The mutual conductance obtained by this means needs only to be reduced by a fraction corresponding to the fraction of the total current intercepted by the screen grid. Typical values of mutual conductance are in the range of 5000–20,000 micromho. As is the case with triodes, the mutual conductance varies approximately as the cube root of the cathode current

The dynamic plate resistance may be estimated from the reciprocal of the slope of the plate-current–plate-voltage characteristics. It is commonly in the range of several hundred thousand ohms to several megohms The dynamic plate resistance tends to vary inversely as the cube root of the cathode current. Because the dynamic plate resistance is so high, the amplification of a pentode amplifier stage can be estimated closely by simply taking the product of the mutual conductance and the load resistance.

The amplification factor is relatively constant and may be estimated by considering that the pentode is really a combination of three triodes

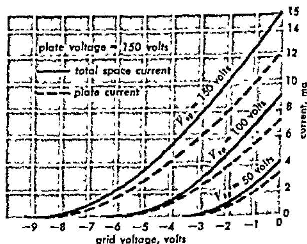


Fig 5. Plate current and space current as a function of grid voltage in a pentode (From K. R Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

found by taking the three possible combinations of three adjacent electrodes. The over-all amplification factor will be the product of the amplification factors of these three hypothetical triodes.

Plate-current cutoff. Pentodes may be classified according to their plate-current–grid voltage characteristics.

Sharp cutoff pentode The so-called sharp-cutoff pentode is the more common form used for ordinary amplifier circuits. The control grid of this tube has a uniform geometrical form over the entire length of the cathode. In a tube with such a grid, the plate-current–control-grid voltage characteristic, as shown in Fig 5, comes sharply to zero as the grid voltage is made more negative. This is the reason for the designation of this type of tube. The sharpness of the cutoff is of course a relative matter; that shown in Fig. 5 is relatively sharp as compared to the type to be described next.

Remote-cutoff pentode. Another type of pentode commonly used is the so-called remote-cutoff pentode. This is also referred to as the variable-mu (μ) pentode. In this type of pentode the control-grid structure is not uniform over the entire length of the cathode. At one end the grid is constructed so that the amplification factor μ is relatively low, and the structure is then tapered to the other end where it is such that the amplification factor is relatively high. This can be done most simply by using a helical grid whose pitch decreases progressively from one end to the other, so that at one end the spacing between wires is large, giving rise to a low μ, whereas at the other end it is small, giving rise to a high μ. With such a tube the current is cut off gradually as the control grid is made more negative. This happens because the high-μ end of the grid cuts off first, as the control-grid voltage is made more negative, and then gradually the entire grid cuts off more and more until the current is finally reduced to zero.

The structure described makes it possible to change the amplification of a pentode stage by changing the grid voltage. This characteristic is caused by the gradual change in mutual conductance with grid voltage over a large range of grid voltage. This tube is useful in automatic gain-control circuits. In these circuits a voltage is developed proportional to the signal strength and is fed back to the grids of some pentodes of the remote-cutoff or variable-mu type in such a direction as to keep the output constant. It is also a consequence of this type of operation that the distortion, or intermodulation, between two signals is reduced. This results because the current characteristics of the pentode do not exhibit sharp curves or corners but have more gradual transitions, a characteristic that contributes to reduced distortion and intermodulation. [K.R.S.]

Penumbra

That portion of a shadow illuminated by only part of a radiating source. A penumbra exists only when the radiating source, usually a light source, has an appreciable angular size. With a point source, the transition from light to dark at the edge of a shadow is abrupt. This is not so, however, with an extended source, where the opaque object forming the shadow may block off the radiation from only a part of the bright surface of the source. Depending on what fraction of this surface is exposed, the illumination in the penumbra varies from zero at the edge of the full shadow to the maximum where the entire source is exposed. *See* SHADOW; *see also* ECLIPSE, ASTRONOMICAL; UMBRA. [F.A.J.]

Pepper

The common garden pepper (*Capsicum annuum*), a warm-season perennial of American origin which belongs to the plant order Tubiflorales. This species includes all the peppers grown in the United States with the exception of the variety Tabasco (*C. frutescens*). Other species, *C. pubescens* and *C. pendulum*, are grown principally in South America. None are related to *Piper nigrum*, the woody plant from which black and white pepper are obtained

Sweet, or nonpungent peppers, generally picked when immature, are commonly cooked or eaten raw in salads Popular varieties are California Wonder and Yolo Wonder. Perfection is the most popular pimento variety grown for canning.

Hot or pungent peppers are most often harvested when ripe and ground into powder for seasoning, although some are canned. The pungent compound, capsaicin, is concentrated mainly in the placental or seed tissue. Popular varieties are Anaheim, Cayenne, and Chili

Propagation is by seed with plants started in greenhouses or outdoor beds and transplanted to the field after 6–10 weeks Field spacing varies; plants 18–24 in. apart in 30–36-in. rows are common Long warm seasons favor high yields and quality; however, high temperatures (above 90°F) and low humidity inhibit normal fruit set.

Harvesting of green sweet peppers begins when the fruit are near full-size but before they become mature and turn red or yellow, usually 60–80 days after field planting. Hot peppers are picked fully ripe for drying, generally 70–90 days after field planting. Florida, Georgia, and New Jersey are important sweet-pepper producing states. The total annual farm value in the United States is approximately $22,000,000 See PAPRIKA; PIMENTO; TUBIFLORALES; VEGETABLE GROWING. [H.J.C.]

Pepper (black)

One of the oldest and most important of the spices It is the dried, unripe fruit of a weak climbing vine, *Piper nigrum*, a member of the pepper family (Piperaceae), and a native of India or Indomalaysia. The fruits are small one-seeded berries which, in ripening, undergo a color change from green to red to yellow. When in the red stage, they are picked, sorted, and dried The dry, wrinkled berries [illegible] ... ke the ... pepper ... from

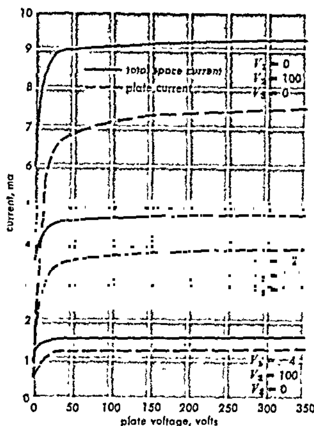


Fig. 3. Plate current-voltage characteristics of pentode. (*From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948*)

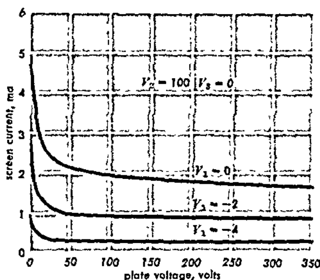


Fig 4. Screen current as a function of plate voltage in a pentode. (*From K R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948*)

the major portion of the cathode current with a certain fraction going to the screen because of direct interception of electrons. The screen current-plate voltage curves all have about the same shape, with the current decreasing as the grid voltage becomes more negative. A final representation of pentode characteristics is shown in Fig. 5. These are the transfer characteristics.

Pentode tube coefficients. The same coefficients that are used for triodes are used for pentodes. These include the amplification factor, the mutual conductance, and the dynamic plate resistance Of these three factors the two most commonly quoted are, in order of their interest, the mutual conductance and the dynamic plate resistance. The amplification factor of pentodes is seldom quoted, because it is so high that it has virtually no significance. The dynamic plate resistance is also high, commonly of the order of megohms, which results in the pentode's exhibiting the characteristics of a constant-current generator. The mutual conductance of a pentode will be similar in its range of values to that of triodes. The significance of the term constant-current generator is that the output current of the pentode will be essentially independent of the load resistance. This assumes what is generally true, namely that the load resistance is small compared to the dynamic plate resistance.

The mutual conductance of a pentode can be estimated by considering that the cathode, control grid, and screen grid constitute a triode to which the triode formulas can be applied. The mutual conductance obtained by this means needs only to be reduced by a fraction corresponding to the fraction of the total current intercepted by the screen grid Typical values of mutual conductance are in the range of 5000–20,000 micromho. As is the case with triodes, the mutual conductance varies approximately as the cube root of the cathode current.

The dynamic plate resistance may be estimated from the reciprocal of the slope of the plate-current–plate-voltage characteristics. It is commonly in the range of several hundred thousand ohms to several megohms. The dynamic plate resistance tends to vary inversely as the cube root of the cathode current. Because the dynamic plate resistance is so high, the amplification of a pentode amplifier stage can be estimated closely by simply taking the product of the mutual conductance and the load resistance.

The amplification factor is relatively constant and may be estimated by considering that the pentode is really a combination of three triodes

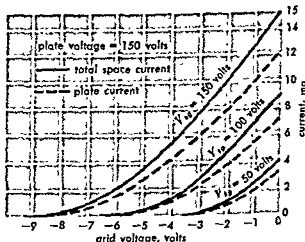


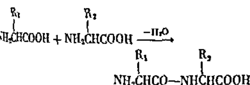
Fig 5. Plate current and space current as a function of grid voltage in a pentode. (*From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948*)

ell as fatigue, infections, and stress Scarring, nstriction, perforation, and hemorrhage are com- on complications. Acute, subacute, and chronic rms of ulcer exist. In some cases, particularly th ulcers of the chronic type, there may be an as- ciation with carcinomatous changes that occur at e base of the ulcer. Other symptoms run the mut of gastrointestinal complaints. See ONCOL- GY [E.G.ST.]

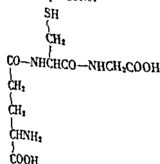
Peptide

A low molec[illegible]

etween the α carboxyl group of one amino acid nd the α amino group of the next:



The peptide is named according to the sequence of amino acid residues in it, beginning with the residue bearing a free α-amino group and moving toward the residue bearing a free α-carboxyl group. If in the figure above, for example, R_1 = H and R_2 = CH_3, the peptide would be named glycylalanine All the residues prior to the terminal one are named as acyl substituents, for example, tyrosylvalylglycylalanine, and the optical configuration of each residue is included when known, for example, L tyrosyl D-valylglycyl L-alanine. Dibasic and dicarboxylic amino acids are not limited to the formation of peptide linkages of the alpha variety. For example, glutamic acid can form γ-peptide linkages and one of the common naturally occurring peptides, glutathione, contains such a linkage, exceptional in natural products:



Synthesis. Most methods for peptide synthesis [in]volve the activation of the carboxyl group so that [th]e carboxyl carbon atom can, in a subsequent step, [re]act with a free amino group Acyl halides, azides and anhydrides are among the active forms that have been employed.

A key problem is that of "masking" the free amino group of the amino acid intended for the N terminal position so that it will not be altered during activation nor react in the subsequent conjugation step. M. Bergmann and L. Zervas in 1932 introduced the use of carbobenzoxychloride (benzyloxycarbonylchloride) as a masking reagent. This widely utilized method is particularly suitable since it does not lead to racemization and because the carbobenzoxy group is easily removed from the peptide by hydrogenolysis under mild conditions. Phthaloylamino acids have also been useful as acylating agents. Here the phthaloyl group is readily removed by treatment with hydrazine

When the peptide is to contain residues of amino acids bearing reactive groups on the side chains, such as cysteine, arginine, and aspartic acid, the problem of preventing side reactions becomes increasingly difficult. The successful synthesis of oxytocin and of vasopressin, octapeptide hormones of the posterior pituitary gland, was an example of inventiveness in solving special problems of peptide synthesis and helped win the 1957 Nobel Prize in Chemistry for V du Vigneaud

Peptides in nature. Glutathione, shown previously, is found in rather high concentration in most plants and animals but its function is not known Carnosine (β-alanyl-L-histidine) and anserine (β-alanyl-1-methyl-L-histidine) have long been known to be present in muscle cells of vertebrates, but again their physiological role remains obscure

The roster of peptide hormones is a rapidly growing one and the following listing shows only those for which a complete sequence has been deduced with the number of amino acid residues indicated in parentheses· oxytocin (8), vasopressin (8), hypertensin (8), melanocyte-stimulating hormone (18), glucagon (29), adrenocorticotropic hormone (39)

Many of the antibiotic substances produced by microorganisms are peptide in nature and, interestingly, many contain amino acids of the unnatural or D configuration. Gramicidin S, elaborated by *Bacillus brevis*, has been shown to be a cyclic decapeptide containing two residues each of L-ornithine, L-valine, L-leucine, L-proline and D-phenylalanine The tyrocidins, the polymyxins, subtilin and bacitracin are further examples of antibiotic peptides See AMINO ACIDS

The capsular material of several microorganisms includes a polyglutamic acid in which the predominant linkage has been found to be of the gamma variety See ANTIBIOTIC; PROTEIN [D.ST.]



Pepper (*Piper nigrum*). (American Spice Trade Assoc.)

the surrounding pulp. *See* PIPERALES; SPICE AND FLAVORING. [P D S.]

Peppermint

This plant, *Mentha piperita*, is an important aromatic herb cultivated in Europe and America. The crop is harvested much as hay is. The dried herbage is hauled to distilleries to remove the peppermint oil by distillation. The leaves are used for flavoring and also brewed to make tea, but the oil is of greatest importance. This is used to flavor candy, gum, and various pharmaceuticals. It is also used as both an external and internal medicine, in soaps, and in perfumes. Menthol, derived from the oil, is a useful antiseptic and is much used in remedies for colds *See* TUBIFLORALES; *see also* SPICE AND FLAVORING. [P.D.S.]



Peppermint (*Mentha piperita*) (USDA)

Pepsin

An enzyme and a constituent of gastric juice, where its function is to aid in protein digestion (*see* ENZYME). It occurs as an inactive material called pepsinogen or pepsin precursor. Pepsinogen is activated by hydrochloric acid, found naturally in the stomach, or by activated pepsin. Thus the reaction is a self-promoting one. *See* PROTEIN METABOLISM.

Pepsin is prepared commercially by extraction from the glandular layer of fresh hog stomach. It is standardized by its activity in digesting coagulated egg white at 52°F where it should handle 3000 times its own weight in 2.5 hours.

Pepsin is part of the crude preparation known as rennet. It is used extensively in the dairy industry in the manufacture of cheese. Milk must be curdled as the first step in cheese manufacture. Pepsin attacks both native or natural proteins and denatured proteins *See* CHEESE; FOOD ENGINEERING; RENNIN. [R E.M]

Bibliography: J. B. Neilands and P. R. Stumpf, *Outlines in Enzyme Chemistry*, 1955; J. B Sumner and G F. Somers, *Chemistry and Methods of Enzymes*, 3d ed., 1953.

Peptic ulcer

A sharply defined ulceration of the upper gastrointesinal tract, characterized by loss of the mucous lining and variable penetration into or through the organ wall. The exact causes are obscure but emotional tension and certain psychological patterns are frequently present in affected individuals.

The lesser curvature of the stomach and the first few inches of the duodenum are the common sites; the lower esophagus and other areas are sometimes involved In most cases the lesion is single and varies in size from a small point to 1 in. or more in diameter. The loss of the mucosa, ordinarily covered by a mucous secretion, lays bare the musculomembranous wall. The crater formed is covered with an exudate lying over the raw tissue, except in chronic ulcers where the pit consists largely of scar tissue.

Most ulcer patients show an increased acidity of the stomach or an abnormal production of acid during times when the stomach is usually quiescent, as at night. Both the action of the acid on raw surfaces and the resulting muscular spasms are thought to be associated with the typical ulcer pain. This pain follows a pattern in most cases, related to the digestive cycle. It appears late at night and may be relieved by food intake or antacids The pain commonly occurs in a specific area and may be precipitated by certain foods or beverages, as